

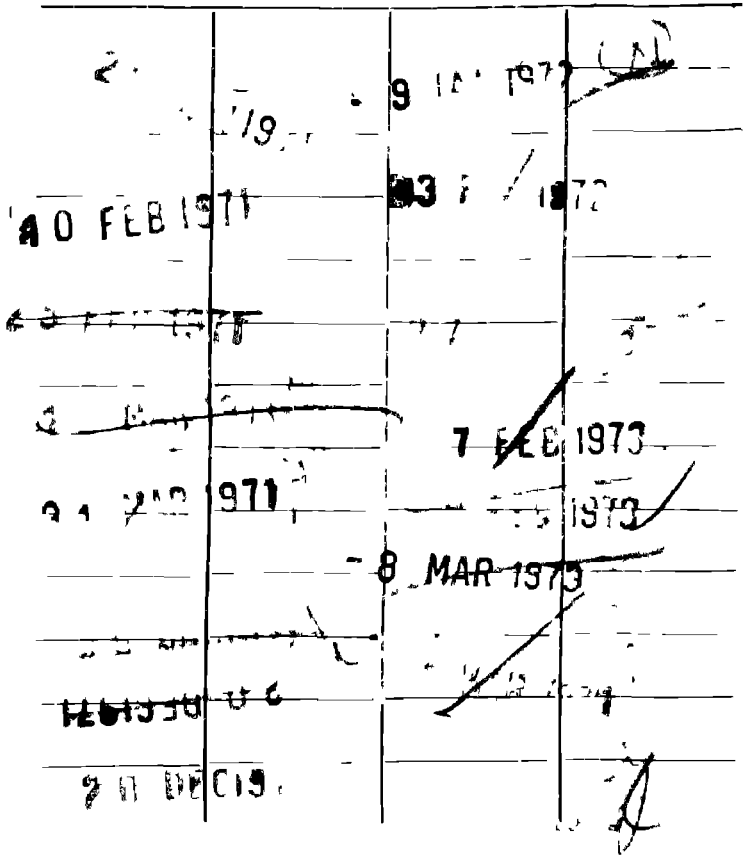
DELHI UNIVERSITY
LIBRARY

DELHI UNIVERSITY LIBRARY

Cl No T:871 H19

Ac No 122891 Date of release for loan

This book should be returned on or before the date last stamped below An overdue charge of 5 Paise will be collected for each day the book is kept overtime



APPRAISING VOCATIONAL FITNESS

APPRAISING VOCATIONAL FITNESS

By Means of Psychological Tests

*

DONALD E SUPER, M A (Oxon), Ph D
Professor of Education, Teachers College, Columbia University
Author of "The Dynamics of Vocational Adjustment"



HARPER & BROTHERS
NEW YORK

APPRAISING VOCATIONAL FITNESS

Copyright, 1949, by Harper & Brothers
Printed in the United States of America

All rights in this book are reserved. No part of the book may be reproduced in any manner whatsoever without written permission except in the case of brief quotations embodied in critical articles and reviews. For information address Harper & Brothers

FIRST EDITION
E-Y

TO MY PARENTS

Margaret Louise

and

Paul Super

*

CONTENTS

	PREFACE	xiii
I.	TESTING AND DIAGNOSIS IN VOCATIONAL GUIDANCE	1
	The Nature and Purposes of Vocational Guidance and Counseling	1
	To Diagnose or Not to Diagnose?	2
	Data Needed in Vocational Diagnosis	6
	Techniques of Gathering Data	7
	The Contribution of Tests to Vocational Diagnosis	11
	Pitfalls in Diagnostic Testing	13
II	TESTING AND PREDICTION IN VOCATIONAL SELECTION	15
	The Peculiarities of Selection Testing	15
	The Importance of Other Techniques	18
	The Validity of Selection Tests	21
	Programs of Testing for Selection, Placement, and Upgrading	23
III	METHODS OF TEST CONSTRUCTION, STANDARDIZATION, AND VALIDATION	28
	Job Analysis	29
	Selection of Traits to be Tested	31
	Selection of the Criteria of Success	32
	Test Construction	43
	Standardization	45
	Validation	48
	Cross-Validation	51
	Factor Analysis and Factor Validation	53
IV	THE NATURE OF APTITUDES AND APTITUDE TESTS	58
	Definitions	58

	The Basic Aptitudes	62
	Methods of Measurement	65
V	Test Administration and Scoring	72
	Arrangements for Test Administration	73
	The Preliminaries to Testing	77
	The Sequence of Tests	78
	Following Directions in Testing	79
	Observing the Behavior of Examinees	81
	Condition of the Examinee	84
	Scoring Tests	85
VI	Intelligence	86
	Nature and Role	86
	Intelligence and Educational Success	87
	Intelligence and Vocational Success	92
	Specific Tests	104
	The Otis Self-Administering Tests of Mental Ability	107
	The American Council on Education Psychological Examination	114
	The Army General Classification Test	124
	The Thurstone Tests of Primary Mental Abilities	132
	The Wechsler-Bellevue Scales of Mental Ability	142
VII	Proficiency	147
	Promise and Proficiency	147
	Educational Achievement Tests	148
	The Iowa Placement Examinations	149
	The Co-operative Achievement Tests	151
	The Tests of General Educational Development	154
	Vocational Proficiency Tests	156
	The Blackstone Typewriting Test	156
	The Blackstone Stenography Test	157
	The Seashore-Bennett Stenographic Proficiency Tests	158
	The Elwell-Fowlkes Bookkeeping Test	159
	Interview Aids and Trade Questions	159
VIII	Clerical Aptitude Perceptual Speed	162
	Is There a Special Clerical Aptitude?	162

Typical Tests 163
The Minnesota Clerical Test 164

IX MANUAL DEXTERITIES 184
Nature and Role 184
Singular or Plural? 184
What is Manual Work? 185
Typical Tests 186
The Minnesota Rate of Manipulation Test 187
The O'Connor Finger and Tweezer Dexterity Tests 203
The Purdue Pegboard 217

X MECHANICAL APTITUDE 221
Nature and Role 221
Specific Tests 224
The Minnesota Mechanical Assembly Test 225
The O'Rourke Mechanical Aptitude Test, Junior Grade 235
The Bennett Test of Mechanical Comprehension 246
The MacQuarrie Test for Mechanical Ability 260
The Purdue Mechanical Adaptability Test 278

XI SPATIAL VISUALIZATION 282
The Minnesota Spatial Relations Test 285
The Minnesota Paper Form Board, Likert Quasha Revision 297

XII AESTHETIC JUDGMENT AND ARTISTIC ABILITY 308
Aesthetic Judgment 309
The Meier Art Judgment Test 310
Creative Artistic Ability 315
The Lewerenz Tests in the Fundamental Abilities of Visual Art 316
The Knauber Art Ability Test 318

XIII MUSICAL TALENTS 319
The Seashore Measure of Musical Talents 321

XIV CUSTOM-BUILT BATTERIES FOR SPECIFIC OCCUPATIONS 330
Business Executives 333

	Dentists	340
	Engineers	341
	Lawyers	343
	Nurses	345
	Pharmacists	347
	Physicians	347
	Pilots	348
	Psychologists	350
	Salesmen	351
	Scientists	353
	Teachers	355
XV	STANDARD BATTERIES WITH NORMS FOR SPECIFIC OCCUPATIONS	358
	The General Aptitude Test Battery	358
	The Differential Aptitude Tests	368
XVI	THE NATURE OF INTERESTS	376
	Definitions	377
	Types of Interests	380
	Socio-Economic Differences	388
	Avocational Differences	389
	Sex Differences	389
	Age Differences	390
	Communality of Interests	392
	Stability of Interests	393
	Family Resemblances	395
	Aptitude as a Source of Interest	397
	Personality and Interests	400
	Origin and Development of Interests	402
XVII	MEASURES OF INTERESTS	407
	The Strong Vocational Interest Blank	408
XVIII	OTHER MEASURES OF INTERESTS	444
	The Kuder Preference Record	445
	The Allport-Vernon Study of Values	465
	The Cleeton Vocational Interest Inventory	471
	The Lee-Thorpe Occupational Interest Inventory	473
	The Michigan Vocabulary Profile Test	474
	Trends in New Measures of Interests	476

XIX PERSONALITY, ATTITUDES, AND TEMPERAMENT 481
Nature and Development 481
Measures of Personality 486
The Bernreuter Personality Inventory 484
The Minnesota Multiphasic Personality Inventory 499
The Bell Adjustment Inventory 510
The Minnesota Personality Scale 513
The Rorschach Inkblot Test 516
The Murray Thematic Apperception Test 524
Trends in the Measurement of Personality 527

XX APPRAISING INDIVIDUAL VOCATIONAL PROMISE 532
Preliminary Considerations 532

XXI USING TEST RESULTS IN COUNSELING 547
Directive Test Interpretation 547
Nondirective Test Interpretation 550
A Synthesis of Suggestions for Interpreting Test Results to Clients 556

XXII PREPARING REPORTS OF TEST RESULTS 564
Reports to Professional Workers 564
Reports to Clients 584

XXIII ILLUSTRATIVE CASES DATA AND COUNSELING 589
Background Data and Test Profiles 590
Counselors' Appraisals and the Immediate Results of Counseling 606

XXIV ILLUSTRATIVE CASES FOLLOW-UP AND EVALUATION 628
The Validity of Vocational Appraisals in the Light of Subsequent Work Histories 628
Conclusions 641

APPENDIX A STATISTICAL CONCEPTS 643
Statistics Quantified Reasoning 643
Central Tendency 644

Dispersion 644
Methods of Expressing Scores 646
The Significance of a Difference 647
Relationship 648

PREDICTION AND PROBABILITY 654
Some Misgivings About Prediction 654
Multiplicity of Factors 655
Taking Contingency Factors Into Account 656
Probability 657
The Accuracy of Estimates 661

APPENDIX B TEST PUBLISHERS AND SCORING SERVICES 665

BIBLIOGRAPHY 667

INDEX 717

PREFACE

VOCATIONAL psychologists are frequently asked, "How good is the Kuder Preference Record (or the Crawford Spatial Relations, or some other test)?" The importance of such questions is brought out by the fact that, in one year, 20,000,000 Americans took a total of 60,000,000 tests (26) Testing is indeed a "big business" It is the aim of this book to provide the user of vocational tests with a detailed and objective answer to questions such as this, for a number of the most widely used and useful tests This is done by bringing together the results of the significant research which has been done with each of these tests, by interpreting these findings in the light of recent developments in testing theory and practice, and by viewing each test in the perspective gained by those who are currently using them in schools, colleges, consultation services, business, and industry

But the objective of this book goes beyond that of providing a manual of currently usable tests, important though that is In bringing together and interpreting the results of research with existing tests, an attempt is made to familiarize the reader with the bibliographical sources and to take him through the processes of collection of data and synthesis of findings, so that he may develop the work habits and thought processes which will enable him, as new research is published and as new tests are put on the market, to evaluate instruments himself and to make new applications Insofar as this goal is accomplished, the user of vocation tests will be enabled to keep abreast of progress in the field and to work on a high professional plane

In this process, the student should develop an understanding of the basic procedures of the development of vocational tests It is true, of course, that most vocational counselors, psychometrists, and personnel workers are and should be primarily users and interpreters rather than constructors of tests It is rare that real skill as test technician and as counselor are combined in one person But, to be an intelligent consumer,

one must be familiar with the procedures and problems involved in the development or manufacture of the product which is to be used This does not necessitate skill in manufacture, but it does require detailed knowledge of methods, materials and problems As each test is studied, the methods used in constructing, standardizing, and validating it will therefore be described in some detail The underlying assumptions will be pointed out, and the validity of the criteria used will be considered Such knowledge is important in personnel selection, in which custom-built tests generally prove most effective, and in vocational counseling, in which generalizations are made on the basis of limited data

As a result, the reader should become well acquainted with the demonstrated values and limitations of the most widely used vocational tests The word *demonstrated* should be emphasized, for during the past twenty or twenty-five years, and especially during the past decade, a great deal of research has been carried on and published on the validity of vocational tests There is no longer any excuse for depending primarily on hunches as to the vocational significance of special aptitude tests, nor for going to the other extreme and concluding that, since "a test tests only what it tests," one can conclude nothing from psychological test results concerning vocational promise Both of these attitudes and practices were widespread during the 1930's, when validity data were sketchy and often disappointing For example, the O'Connor Tweezer Dexterity Test was frequently used as one indicator for dental training, on the basis of the test author's unsupported statement that it should be valid for dentistry, and on the basis of logical analysis Some counselors and personnel workers, however, impressed by the lack of expected validity in some of the tests for which criterion data had been obtained, refused to concede any predictive value to tests, maintaining that aptitudes are too highly specific for performance on one laboratory task to predict performance in a real life situation Enough data have now been accumulated so that a more realistic and pragmatic approach is possible the counselor can know, from experimental evidence, a good deal about the nature of the trait being measured and about its rôle in vocational adjustment His interpretations of test results can therefore be based on objective evidence or, when the evidence does not go far enough, on logical analysis which uses fact rather than fancy as a starting point

It is not meant to imply, however, that we now know all we need to know about aptitudes and interests, nor about the instruments which we use to measure them On the contrary, there are still many gaps in our

knowledge, some of them surprising indeed after a generation of creative work. For example, such a simple question as that of the maturation of clerical aptitude as measured by the Minnesota Vocational Test for Clerical Workers (speed and accuracy of name and number discrimination) has not been answered, despite some beginnings, or, putting it in practical rather than theoretical terms, we do not yet know at what age it is legitimate to use adult norms for the Minnesota Clerical Test, and at what ages comparison should be made only with boys or girls of the same age level The question of the relationship between two and three-dimensional spatial visualization has not yet been finally answered, fundamental though it is to the use of the Minnesota Spatial Relations and Paper Form Board Tests in shop work as opposed to drafting Even apart from somewhat theoretical questions there is still much to be done The norms for one of the most valuable group tests of intelligence, the American Council on Education Psychological Examination, for example, are still entirely based on college freshmen, research has shown that scores increase with age in college (see p 115), but we have practically nothing concerning the occupational significance of A C E scores at any age, something that it would seem both logical and important to have for use in counseling college students This point is dwelt upon briefly, partly in order to stress the fact that, although we know a great deal about the significance of many tests, there are still great gaps in our knowledge, and partly in the hope that the pointing out of some of these gaps will result in further research along lines which will round out our knowledge

One of the principal weaknesses in the measurement movement has been the excessive individualism of the research which has been carried on Individualism has been good in that it has encouraged branching out in new directions and trying out new possibilities, but it has been bad in that it has resulted in the scattering of efforts and in the frequent dropping of a good idea after it has been barely tried For every research project comparable to Strong's persistent study and refinement of his Vocational Interest Blank throughout the past twenty years, there are several like Zyve's Scientific Aptitude Test and Bernreuter's Personality Inventory, whose initial promise have never been adequately explored. This is partly because the test authors, often for excellent reasons, did not follow up their initial work, partly because the research carried on by other people with these instruments has generally been unco-ordinated and incidental

For test development work to be fully effective, two things are needed

in addition to those which have characterized it so far One of these is the periodic and systematic review of work with specific tests or types of tests This should be more detailed, critical and creative than the periodic reviews published in the *Review of Educational Research* by the American Educational Research Association, it should be more regular and more co-ordinated than the excellent reviews which occasionally appear in the *Psychological Bulletin* and *Psychological Review* as a result of the activities of individual psychologists, and it should be more integrated and pointed toward action than Buros' *Mental Measurement Yearbooks* (126)

It is hoped that this book will serve this purpose, pointing out important research that needs to be done to round out our knowledge of vocational tests and stimulating psychologists, vocational counselors and personnel workers to carry out appropriate research projects There should in time be a committee of the American Psychological Association, the National Vocational Guidance Association, and the American Management Association whose function it is to plan and co-ordinate such critical and constructive reviews The second major need in the development of vocational testing is an extension of this function from systematic review and suggestion to systematic planning and execution of research Such a committee should take the initiative in encouraging research along needed lines, partly by publications and talks at professional meetings, and partly by a program of grants-in-aid of suitable research Assistance should even be provided in planning and financing major research projects for the large-scale study of a number of important and related problems The Minnesota Mechanical Abilities Project of the 1920's, the Minnesota Employment Stabilization Research Institute of the 1930's, Strong's work in vocational interests, Thurstone's work on primary mental abilities, Kuder's work on primary interests, the United States Employment Service's work on the development of basic occupational test batteries should be multiplied and, in some cases, expedited as they could be only in a nationally sponsored and co-ordinated plan

A few words should be said about the selection of tests discussed in this book No attempt is made to cover all tests, or even all tests of some value Annotated catalogues of tests are available from publishers and distributors such as the Psychological Corporation, Science Research Associates, World Book Company, and California Test Bureau Too many treatises of testing are little more than annotated catalogues Instead, a number of tests have been selected for detailed consideration because they measure aptitudes or traits of demonstrated importance, are typical of

others designed to measure the same characteristics, are as readily administered and scored as others of their type, and, particularly, have been sufficiently studied so that something is known both concerning the nature of the characteristics measured and the validity and usefulness of the measuring instrument In a few instances this last, most fundamental, consideration has been departed from in order to permit brief discussion of what appears to be a promising technique deserving of more extensive and thorough study In addition, briefer mention is made of certain other tests which merit discussion because they are widely known even though of little or no vocational value Discussion of some tests seems forced upon one by the extent of their use in industry, even though neither their proved nor probable value to the counselor or personnel man justifies giving them space Similarly, the use of Wechsler-Bellevue part-scores as indices of special aptitudes by many clinical psychologists dealing with problems of vocational adjustment makes it necessary to consider that topic, even though there is as yet little occupational evidence to justify such a practice The tests discussed include all but six of the 40 tests listed by Berkshire et al (83) as most commonly used in guidance centers, plus several less widely used but otherwise important instruments These authors state that some 20 of the tests surveyed appear to be "basic to the guidance function" Similarly, the great majority of tests found, in a confidential survey of industrial testing, to be widely used are included in this treatise

Apart from the annotated catalogue approach which has characterized a number of books on testing, several other approaches are possible One of these is the introductory survey of measurement theory and practice E B Greene's *Measurements of Human Behavior* (309) is one of the most widely used examples This present book differs from such texts in that it assumes a knowledge of the fundamentals of measurement (of which a review is provided in Appendix A for those who need it), and in that it deals with the problems, methods, and results of vocational testing in an intensive and comprehensive manner It is designed to serve both as a handbook for counselors, psychometrists, and personnel workers actually using tests in practice, and as a text for courses in the use of tests in counseling and selection

Another approach in a book or course on testing is to teach the techniques of test construction and validation Clark Hull's *Aptitude Testing* (385), a classic in this field for more than a decade after its publication in the mid-twenties, illustrates this emphasis Adkins' *Construction and*

Analysis of Achievement Tests (7) is a more recent manual, written for personnel selection Thorndike's *Personnel Selection* (833a) is another There is a need for a text of this type, for use in courses on test construction, but this book does not attempt to meet both needs

Still another approach is that embodied in Walter V Bingham's *Aptitudes and Aptitude Testing* (94), published under the aegis of the National Occupational Conference in 1937, for a decade the standard text in courses on testing in vocational guidance, and now undergoing revision In his book, Bingham focuses attention on the constellations of abilities that play a part in success in the major occupational fields This occupational orientation is important, but in stressing it, something more important to the user of tests in actually understanding a person who has been tested is neglected This is the consideration of the question, *"what does this test, and the score made on it by this person, tell me about his vocational promise?"*

It is around this question that the author has attempted to organize this book Experience as counselor, personnel consultant, supervisor and instructor has shown that the user of vocational tests in diagnostic work starts with data about the *client,* which he then synthesizes and interprets in terms of *vocations* It is true that he needs to make a decision as to what vocational goals are likely to be considered in order to select appropriate tests, and that this requires thinking in terms of occupations and constellations of abilities However, test batteries for occupational families are not yet developed to a sufficient degree to make this the best approach in actually interpreting test results and counseling Instead, the psychologist or counselor must tease what meaning, suggestions, and contra-indications he can from the test and other personal data on hand In some of the most effective vocational counseling and personnel evaluation services vocational tests are used, not only for the occupational norms which permit comparison with successful workers, but also for the analysis of the psychological strengths and weaknesses of the client, which are then interpreted in terms of possible vocational opportunities This latter type of analysis requires thorough knowledge of the tests used, supplemented by detailed knowledge of occupations from first-hand experience and from psychological research This book therefore considers the topic stressed by Bingham, but emphasizes that which he played down, in the belief that this is more helpful to the user of tests Another unique feature in a text such as this is the material on the use of test results in counseling, that is, on putting test results to work

It is the writer's belief that this book should be of special value to vocational psychologists, personnel workers, and counselors in another way Great progress has been made in testing for vocational selection and guidance (the two go hand-in-hand) during the past ten years Much of this work has been published in the journals and monographs, much is still in the files of the military services, some is simply part of the folklore of vocational testing and counseling, known to some of those engaged in such work The writer hopes that, in drawing from intimate knowledge of these several sources, he has been able to make the most important of these advances available to users of psychological tests in vocational guidance and selection If the work of the Aviation Psychology Program of the Army Air Forces has been drawn on more extensively than any other single source, it is because the comprehensiveness and thoroughness of that program made it a unique source of materials on personnel testing

In using this book as a text in a graduate course in vocational testing, the author uses four other instructional aids which may be of interest to other instructors Although they have been developed to supplement the book, they are independent of it as it is of them One of these aids is a *Sourcebook for Vocational Testing* (Teachers College Bureau of Publications), containing photo-offset reproductions of a number of the more significant original articles on the tests dealt with in this book, it is used to facilitate access to journal material and to train students to use reports of original research in evaluating and understanding tests The second is a *Kit of Vocational Tests* assembled by the Psychological Corporation and the College Bookstore, it contains manuals, scoring keys, and test blanks for all paper and pencil tests studied intensively in the course (the major tests treated in this book) Students thus have easy access to manuals and keys, and start their own test libraries The third aid consists of copies of catalogues of selected test publishers, giving complete data on ordering and costs, this makes it unnecessary to include such transitory data in the text The fourth aid consists of a well-equipped testing laboratory, in which supervised practice in testing is given

It is with mixed feelings that the author parts with this manuscript Based as it is on the findings of research in a rapidly developing field, it is inevitable that even before it comes off the press some of the questions which have been mentioned as unanswered will have been answered by new investigations Some of the conclusions may soon need modification The indulgence of the reader is therefore requested when he finds that

the facts on which a generalization is based have changed The material in this book should nevertheless be of vital importance, if only as a background against which to see the findings of new studies as they appear It is the writer's intention to revise it periodically, as new tests and new findings require He will therefore welcome the co-operation of authors of research studies in sending him reprints of papers bearing on the subject of this book By cutting down the time consumed in bibliographical research, this will make it easier to survey relevant material, improve the coverage of the book, and speed up the preparation of revisions

The acknowledgments due to others in connection with the preparation of this book are numerous, varied, and a source of such pleasure that I have looked forward to the writing of these paragraphs

First, there are those from whose work I have learned much of what I know about testing Professor Donald G Paterson, Dean Edmund G Williamson, and Dr John G Darley, of the University of Minnesota, the first-named an unseen friend whose correspondence over a period of several years has added to the professional stimulation provided by the publications of the Minnesota researchers, Dr Edward K Strong, Jr , of Stanford University, whose work in the measurement of interest first aroused my interest in measurement, Drs Laurance F Shaffer, Neal E Miller, and Robert R Blake, at one time officially and respectively my chief, associate, and assistant in the Aviation Psychology Program of the Army Air Forces, but actually my helpful and stimulating colleagues in a number of research projects, Dr John C Flanagan, now of the American Institute for Research, formerly director of the Aviation Psychology Program of the Army Air Forces, whose vision and singleness of purpose made that program both a landmark in the field of psychometrics and a most worth while professional experience for those involved in it, and Dr Harry D Kitson, my senior colleague, whose interest in improving the understanding of vocational tests by their consumers has been a constant encouragement in the preparation of this book

Secondly, there are those who have contributed to the actual writing of the book by their careful reading and criticism of parts of the manuscript Dr Kitson read the first draft in its entirety, applying his skill and perspective as editor of *Occupations* to the broader problems of organization, presentation, and interpretation Professor Paterson read selected chapters of which his experience in test construction and perspective as editor of the *Journal of Applied Psychology* made him a valued critic Dr Shaffer also found time in his busy schedule as Chairman of the Department of

Guidance, Teachers College, Columbia University, and editor of the *Journal of Consulting Psychology*, to read the introductory chapters, the chapters on tests of intelligence and of personality, and those on the use of test results, with unusual care and discernment Mr Bruce Shear, Director of Pupil Personnel Services for Northern Westchester County, has made practical suggestions concerning certain chapters Charles N Morris, my junior colleague, Stewart Murray, Director of Guidance for Nova Scotia, Vernon Wallace, Counselor at Brooklyn College, Davis Johnson, Counselor in the Vocational Counseling Service of New Haven, Joseph B Shay, Psychologist in the Jewish Vocational Service of Detroit, and David Lane, Associate Director of the Veterans' Guidance Service, Clark University, read parts of the manuscript as graduate students, checking many details, pointing out professorial obscurities, and encouraging me with their constant interest

Thirdly, there are the authors and publishers who have graciously made possible quotation from their works, particularly to the American Book Co , the American Psychological Association, Henry Holt and Co , the Houghton Mifflin Co , Dr G Frederic Kuder, the McGraw-Hill Book Co , *Occupations*, the Psychological Corporation, the Science Research Associates, the Social Science Research Council, and the Stanford University Press In addition, Dr Harold G Seashore of the Psychological Corporation and Mr John R Yale of the Science Research Associates cooperated in supplying data and checking facts concerning certain tests

The final word has been saved for the women and the children Miss Esther Grossmark, my secretary, has with patience and persistence supervised part-time typists and sandwiched the typing of parts of the manuscript into a heavy workload, strengthened, no doubt, by the special interest of a student of psychology And my wife and sons have cheerfully spent innumerable weekends and evenings in other parts of the house and garden while the typewriter hammered away in the study, breaking the monotony occasionally with a pleasant word or an excited account of some neighborhood event

DONALD E SUPER

Montclair, N J
February, 1949

APPRAISING VOCATIONAL FITNESS

CHAPTER I

TESTING AND DIAGNOSIS IN VOCATIONAL COUNSELING

The Nature and Purposes of Vocational Guidance and Counseling

VOCATIONAL counseling has two fundamental purposes to help people make good vocational adjustments and to facilitate the smooth functioning of the social economy through the effective use of manpower

These purposes imply that each individual has certain abilities, interests, personality traits, and other characteristics which, if he knows what they are and how they may be turned into assets, will make him a happier man, a more effective worker, and a more useful citizen Part of his education, that is, literally, "leading him out" or guiding his development and unfolding, therefore consists of helping him to get a better understanding of his aptitudes for acquiring various skills, his adaptability to differing types of situations, and his interest in the numerous activities in which he might engage Although less generally recognized as such, this self-understanding is just as much an objective of education as is the development of an understanding of the world in which he lives A well-educated man is one who has achieved both types of understanding, a well-adjusted man is one who has been able to put these two types of knowledge to good use and has found a place for himself in society

Some educational programs have assumed that the processes of mental discipline, intellectual development, and general education would result in the desired self-understanding However legitimate this assumption might be in an effective educational program, the result is not achieved in practice the Regents Inquiry into the Character and Cost of Public Education in New York State, as reported in the monographs by Eckert and Marshall (234) and by Spaulding (729), made it clear that a large proportion of the products of our more or less traditional school systems have neither the self-understanding nor the understanding of the world around them that is necessary for good vocational adjustment or citizenship This lack of self-insight and of social understanding has been re-

vealed by numerous other studies of the relationship of the vocational aspirations of youth to their abilities and to the opportunities open to them (793 Ch 2)

This being the case, vocational guidance is needed, to focus attention on the information about self and occupations that is needed for good vocational adjustment and to guide the development of a genuine understanding and acceptance of these facts Vocational guidance is, therefore, a dual process of helping the individual to understand and accept himself, and of helping him to understand and adjust to society, it is both psychological and socio-economic

What are the psychological processes necessary to bring about the understanding which experience alone so often fails to produce? They are, of course, those of vocational counseling And what is vocational counseling? It is the process of helping the individual to ascertain, accept, understand, and apply the relevant facts about himself to the pertinent facts about the occupational world which are ascertained through incidental and planned exploratory activities The techniques of vocational counseling vary from case to case and from counselor to counselor, depending partly upon the counselee's state of readiness and partly upon the time available to the counselor, the degree of skill he has attained, and his philosophy of counseling In many cases these techniques fall naturally into two categories those of diagnosis and those of treatment or counseling in the more limited sense There is, however, one important school of thought in guidance which is sometimes described as opposed to the use of diagnostic activities, at least of the traditional varieties and in the traditional ways This point of view has been most ably and widely propounded by Carl Rogers and his students (639,640,641) and is known as nondirective counseling Before embarking upon a discussion of the techniques of diagnosis prefatory to the intensive study of diagnosis through tests, some consideration should be given to this question of the role of diagnosis in vocational and educational counseling

To Diagnose or Not to Diagnose?

Nondirective counseling is based on the assumption that the individual has, within himself, the resources necessary to the solution of his own problems All that he needs, according to this theory, is a permissive situation, one in which he can release his energies and bring these resources into play It is the counselor's role to create this permissive

situation and to release these energies He does this by creating a warm and understanding atmosphere, by accepting and reflecting the feelings of the client, and thus making it possible for the client to work out his problem in his own way

Nondirective counseling originated in the treatment of behavior problems by child guidance workers such as Jessie Taft working under the leadership of Otto Rank, and was referred to by them as passive, as contrasted with active, relationship therapy It was further developed by Rogers in working with the more normal personality problems of adolescents and adults as well as with children's behavior problems (638,639,641), he and his students did a great deal to clarify the principles, systematize the procedures, and broaden the applications of passive relationship therapy in research and in teaching as well as in clinical work, in the process it was renamed nondirective therapy Having demonstrated the values of nondirective counseling in dealing with certain types of personal adjustment problems, Rogers and some of his students have moved on to consider its application to problems of vocational and educational counseling (166,173,640)

Having worked primarily in clinics and with the mild and moderate neurotics who turn to psychological clinics for help in quite disproportionate numbers, Rogers has been impressed by the number of presumed problems of vocational adjustment which turn out to be problems of personality adjustment "For the nondirective counselor, vocational and educational difficulties *are* personal problems" "Following the viewpoint of this manual will usually demonstrate that the statement of a vocational or educational problem really disguises a deeper personal problem that must be handled before any real progress can be made on the manifest difficulty" (641 90 and 104) [1] If Rogers and his students had worked in more normal situations, with a more typical sample of adolescents and young adults, they would have found that a larger percentage need vocational guidance but have no significant personality problems and are ready for the "progress on the manifest difficulty," for which, as Rogers states, neurotic clients are ready only after psychotherapy The average high school pupil and college student does not need this (322) Indeed, one of Rogers' students who works in a university guidance center reports that nondirective counseling seems appropriate in about twenty percent of the cases seen in that center (Arthur Combs,

[1] By permission from *Counseling with Returned Servicemen*, by C R Rogers and J L Wallen, Copyrighted 1945, Houghton-Mifflin Co

in an address at the 1946 Regional Conference of the Council of Guidance and Personnel Associations, Hotel Pennsylvania, New York) More research needs to be carried out on this question before a definite conclusion can be drawn, but the evidence so far suggests that what Rogers has demonstrated with clinic cases cannot be applied without modification to school, college, and normal adult cases

This being so, Rogers' injunctions against diagnosis (e g , 641 5–6) cannot be lifted from his discussions of psychotherapy and applied to vocational counseling This is not the place to dwell upon the adequacy of Rogers' views on the wisdom of avoiding the diagnosis of personality problems (see Patterson, 594), although it might be pointed out in passing that he does advocate some diagnosis when he writes (641 104) "The meaning of the personal relationship must be *assessed* [italics mine] What use is the client attempting to make of his relationship with the counselor?" More important here is the fact that he states, in discussing a case (641 94), "True, the information was important in helping him to evaluate himself more realistically than he had previously, but only because the counselor allowed him to work through his attitudes and feelings about the situation in the light of the new information" In other words, diagnosis skillfully done, at the right stage, and integrated with the counseling, is often desirable

As the writer sees it, Rogers' sketchily expressed and scattered views on diagnosis in vocational counseling amount to this many cases which seem to be problems of vocational and educational counseling are in reality personality problems, and therefore it is wise to use nondirective techniques at least in the first contact in order to establish the nature of the real problem, if or when the real problem is vocational or educational, the diagnostic use of tests may provide needed and valuable information concerning the client which he will want to take into account in making his plans, when such information is obtained and used, its emotional significance to the client needs to be worked out by nondirective methods, especially if the client is also working through problems of personality adjustment

Bragdon (116 81) and Fisher and Hanna (257) have reported in early studies, and the writer has pointed out in his text on vocational guidance (793 205,207,215), that many problems which appear to be vocational and educational are in reality personal, this has been a widely accepted fact among vocational counselors The evaluation of the client's reaction

to diagnostic data shared with him, combined with discussion (sometimes nondirective and sometimes rather directive in nature) designed to help him understand and accept the facts is similarly an old and widely-used technique of vocational counseling, as one who has observed many mature and experienced vocational counselors at work can testify Rogers' contribution seems to have been to stress these facts in a way which have brought them to the attention of other counselors who have been more directive in their approach and who have tended to emphasize their own diagnostic activities at the expense of the client's understanding

If Rogers' views are contrasted with those expressed by Williamson in his book on counseling (928 133–142) the error into which those who rely too much upon tests, or are primarily interested in problems of diagnosis, too easily fall will become clear The type of counseling outlined therein is quite directive, as Darley expresses the same point of view in another book (190 169), "the interview seems somewhat similar to a sales situation, since the counselor attempts to sell the student certain ideas about himself, certain plans of action, or certain desirable changes in attitudes "[2] The assumption is that since the counselor obtains the significant information by technical methods and is better qualified to understand their significance than the counselee, he should seek to convey the information to the client by rational means and to get him to adopt an appropriate plan of action To quote Williamson (928 136) "Ordinarily the counselor states his point of view with *definiteness,* attempting through exposition to *enlighten* the student "[3] Williamson's fallacy, like that of many who have been concerned more with the development of diagnostic techniques than with the development of individuals, seems to have been to expect the counselee to gain insight by the same rational processes used by the counselor in making a diagnosis[4] As many other counselors have long known, and as Rogers has very effectively reminded us, the insight-gaining processes of the counselee are affective and not cognitive, they are emotional rather than rational When objective evidence is shared with the client his subjective reactions to it need to be aired and examined in a way peculiarly suited to nondirective inter-

[2] By permission from *Testing and Counseling in the High School Guidance Program,* by J G Darley Copyrighted 1943, Science Research Associates

[3] By permission from *How to Counsel Students,* by E G Williamson, Copyrighted 1939, McGraw-Hill Book Co

[4] The reader may wish to refer to the original context as Darley has indicated the belief that such quotations do not adequately represent this view see *J Appl Psychol,* 1944, 28, 179–180

viewing If this type of diagnostic activity is carried out well, progress in vocational adjustment will be facilitated

Data Needed in Vocational Diagnosis

In order to evaluate a person's vocational prospects, two types of information about him are needed the *psychological* facts which describe his aptitudes, skills, interests, and personality traits, and the *social* facts which describe the environment in which he lives, the influences which are affecting him, and the resources which he has at his disposal To depend upon one type of fact to the neglect of the other is to be unrealistic and to disregard important elements in vocational adjustment, for the opportunities available to persons with similar aptitudes and interests may vary greatly, just as the abilities and traits of people in similar social situations differ from one person to the next It has, for instance, been demonstrated that many young men and women capable of benefitting from a college education do not attend college because of financial handicaps (234), just as many students who can afford to attend college drop out because of learning difficulties

The fact that many psychological characteristics are best judged by means of tests which require special study and have the appearance of objectivity and concreteness has often led to the relative neglect of social factors in counseling by those trained to use tests, and to the neglect of important psychological factors by those not trained to use tests For these reasons it seems desirable, in considering the types of data needed in vocational diagnosis, to stress the need to obtain both types of information and to use both testing and non-testing techniques More will be said later about the methods of gathering data, first let us focus on the types of data needed

Psychological data needed include information concerning the general intelligence of the individual, that is, his ability to comprehend and use symbols or to do abstract thinking This academic aptitude is important not only in school situations, but also in everyday life situations in which ability to analyze a situation or a problem, to draw conclusions, to generalize, and to plan accordingly, is needed Special aptitudes must also be explored The work of recent years has shown that what has been thought of as general intelligence is, in reality, a combination of special aptitudes such as verbal comprehension, arithmetic reasoning, and spatial ability (281) For this reason data concerning strength or weakness in any one of these special areas must be obtained Other special

aptitudes which play a part in clerical, technical, musical, artistic, and manual activities must be known. The subject's interests, attitudes, and personality traits need to be assessed, in terms of their vocational implications. And finally, data are needed as to the degree of proficiency which he has attained in using any of the skills which he has acquired

Social data are needed in order to provide a framework in which to interpret the psychological data The occupational level of the parents plays an important part, for example, in determining the vocational ambitions of a youth and in his drive to achieve them, as well as in fixing the financial resources upon which he can draw in furthering his ambitions The vocational achievements of the subject's brothers and sisters may be indicative of his own probable level of achievement, but this prognosis is modified, in turn, by the age of the parents and their financial independence It not infrequently happens that the youngest child fails to reach an occupational level as high as that of his siblings because of the need to contribute to his parents' support just at the time at which he might have been going to college The industrial and cultural resources of the home and of the community, the educational experiences of the individual, his leisure-time activities, and his vocational experiences all need to be examined, in order that the resources open to him and the use he has made of them may be understood To draw the line between psychological and social data is obviously impossible at times, for in finding out what influences have been at work on a person one also ascertains the ways in which he has reacted to them

Techniques of Gathering Data

With the improvement of testing techniques it has become possible to measure an increasing number and variety of important psychological characteristics In 1918 intelligence was the only psychological characteristic of vocational significance which could be effectively measured, in 1928 manual, mechanical, artistic, musical and spatial aptitudes, and vocational interest, could be added to the list, although the measures of these characteristics were then quite new and therefore relatively little understood By 1938 a considerable amount of information had been gathered about and by means of these instruments, they had been refined and improved, and attitudes and clerical aptitude had been added to the list of measurable entities In 1948, after the lapse of another decade, further improvements have been made in existing types of instruments, much more is known about them, and measures of personality have been

developed to a point at which they appear to have clinical validity even though their vocational significance is not clear

Despite the great progress in psychological testing since World War I, the variety of characteristics which can be measured still leaves a great deal to be desired As is made clear in greater detail in subsequent chapters, the measuring instruments we now use even for the most adequately measured traits such as intelligence and vocational interest are still crude and only half-understood, those we use for measuring personality traits such as general adjustment, introversion and the need for recognition are still in embryonic stages, and there are no methods of testing creative imagination, persistence, and certain other traits and abilities which are often assumed to be important and which laboratory studies and other types of investigations have suggested may actually exist

For these reasons the psychological study of a person's abilities and personality traits requires more than testing techniques When a suitable test is available, its use will generally save time and obtain the information in a more objective, valid, and usable form than would otherwise be the case This is especially true of intelligence, and it applies also to a variety of other traits But some tests measure aspects of ability or interest which are so narrow as to make their use dangerously misleading unless the data obtained with them are thought of as being only one small part of the aptitude picture, for example, the existing tests of musical talent do not measure anything as broad as that term implies but only certain minute aspects of musical aptitude They need to be supplemented by observation of musical performance, ratings by musicians, history of interest in musical activities, etc As the major part of this book is devoted to the uses of vocational tests, it is the purpose of this section to point out some things that tests can*not* now do rather than to show ways in which they are useful It aims to indicate briefly the nontesting techniques which must be used in order to obtain a well-rounded picture of a subject, rather than to discuss the useful testing techniques

The *interview* is the most widely used subjective method of gathering personal data, as well as the principal treatment or counseling technique In diagnosis as in counseling, there are traditionally two divergent points of view concerning interviewing In one approach, the emphasis is on careful planning, in having a well-thought-out interview schedule or form which is to be completed during the interview The interviewer asks direct questions, using the phraseology of his schedule and adhering

to the order in which the questions appear on the schedule In the other, the nondirective approach, the interviewer merely sets the topic ("structures the situation"), then accepts and reflects feeling in order to let the person being interviewed lead the discussion into the areas which are most important to him Although the interviewer may not gather data on exactly the topics which he had considered important, he does obtain material on the problems which are of most importance to the client, and therefore most important for diagnosis The Hawthorne Study well illustrates the development of this technique (637 Ch 13) A commonly used procedure is the patterned or semi-structured interview, in which the interviewer uses the schedule only as a guide In this semidirective type of diagnostic interviewing, the essence of the technique is to use key questions as a means of getting the person being interviewed to talk freely on important topics, in the anticipation that desired facts will be brought up in a context which makes their interpretation more complete than it would be if the facts were given briefly and in response to a direct question In either type of data-gathering interview, and especially in the less directive type, it is possible to obtain information not only on factual items such as those normally covered in the social history, but also on attitudes, ambitions, and other affective matters which constitute the psychological case history (see 96, and 768 Ch 3 and 4, for detailed discussions)

Questionnaires are frequently used in order to obtain data such as are commonly gathered in the interview The writer has demonstrated that with literate subjects who want to co-operate this is an effective timesaver in collecting factual material (804), but it is much less useful than the interview as a means of gaining insight into the attitudes and feelings of any but the most frank and insightful of individuals Research by Landis (451) and others has shown that factual items are generally reported with considerable accuracy when the subject has come for counseling, although there is evidence (656) that others, whether subjected to diagnosis against their will or under scrutiny as applicants for positions, yield to the pressure to falsify facts and improve appearances as much as they consider possible Useful material on attitudes can sometimes be gathered by questionnaire methods, often by transforming the questionnaire into an attitude scale, but Spencer (733) has shown that the truthfulness of material obtained depends on the anonymity of the response and, by inference, on the confidence of the respondent in the person using the data Symonds (810 Ch 4) has discussed the details of

questionnaire construction at some length, pointing out steps which can be taken to improve the understanding of the questions by the various people filling out the form and thus to compensate as much as possible for the lack of flexibility inherent in the technique If the questionnaire is well constructed and good rapport is established in its use, remarkably frank answers can be obtained concerning matters which the respondent is able to put into words, as shown in a study made under conditions of anonymity by Shaffer (710) and in another involving signed questionnaires by Kemble (420)

3 *Rating scales* are a third widely used non-testing technique of gathering diagnostic data, although they resemble tests in that they attempt to quantify evidence and to be objective A great deal of research, summarized by Symonds (810 Ch 3) and from the counselor's point of view by Strang (768 Ch 6), has demonstrated that despite its objective appearance the rating scale is a very subjective technique, being fundamentally the recording of opinion Despite this defect, rating scales have been found useful in personnel selection (115) and evaluation (538 195–197), but judging by the accumulated experience of those who have tried them, they have not proven very helpful to counselors interested in getting a picture of the characteristics of students or others with whom they are working

4 *Anecdotal records* resemble some aspects of the better types of rating scales in that they call for descriptions of behavior as observed in concrete situations The American Council on Education Personality Report, for instance, calls for a specific illustration of every characteristic rated on the graphic scale if the student has shown evidence of leadership, the rater is asked to describe a situation in which this was demonstrated An anecdotal record differs, in that it consists of a collection of such incidents described soon after the event and accumulated in the subject's file If the incidents are well chosen and well described (neither of these desiderata can be taken for granted) it is then possible to analyze these records and construct a dynamic and characteristic picture of the individual in question and to make judgments concerning his probable behavior in other situations This technique has been studied by Jarvie and Ellingson (398), and is described by Strang (768 Ch 5) and Traxler (860 Ch 7)

5 *Personnel records* are another source of diagnostic data available to schools, colleges, and business enterprises The data included therein are often so sketchy as to shed little or no light on the abilities, interests,

personality traits, background, or family situation of the person in question, on the other hand, they frequently include a variety of important diagnostic data In a school or college the student's courses and grades are at least likely to be available, while in an industrial concern his type and amount of education, previous employment history, marital status, earnings, and attendance are likely to be on record The case histories of social agencies and credit ratings often provide other material If the records go into more detail concerning the subject's special achievements and problems the counselor or personnel worker has at his disposal data on proficiency, interests, and personality traits which have the advantage of having been accumulated over a period of time and therefore of showing trends of development, and which generally reflect the judgment of a variety of people The principal problem in using personnel records is to keep them sufficiently complete without making record keeping take time that is needed for diagnosis and counseling Strang (768 Ch 2) discusses the use of personnel records in schools and colleges, treatment of their business and industrial uses will be found in Scott and others (685 Ch 8–10) and in Moore (538 Ch 4)

Essays and *autobiographies* provide another source of diagnostic data Counselors and admissions officers in schools and colleges frequently ask students to write an autobiographical sketch, often focussing on their educational and vocational experiences and plans, in order to get an understanding of their interests and motivation There has been much less systematic study of this technique than of most, despite its widespread use It is used not only in educational institutions, but also by foundations granting fellowships, rarely by business enterprises It is briefly discussed by Strang (768 113–116) and at somewhat greater length by Fryer (277 371–419)

The Contribution of Tests to Vocational Diagnosis What has just been said should make it clear that psychological tests are only one way of obtaining information needed to understand a person whom one is counseling To put it concretely, the intelligence of a young man two years out of high school can be judged by an intelligence test administered to him especially for that purpose, by his marks in high school, by his father's occupation, by his own occupational experience since leaving school, and by various other indices

It is true that all of these methods have defects the test may not truly represent his mental ability because of a reading handicap, his high school marks may not be a good index because of his poor motivation

at that time, his father's occupation may be the result of social stratification rather than of his own enterprise and ability in a fluid society, and his own occupational experience may have been distorted by depression conditions But they also have their own peculiar advantages the young man's occupational history shows what he has actually done with his ability in a situation in which the economic factors are known, there is a demonstrated relationship between intelligence and occupational level, whether the occupation referred to is that of the father or of the person in question, high school marks correlate to a moderate extent with intelligence tests and with subsequent achievement in college, and good tests well given are relatively free from extraneous influences and do yield a prediction of performance or satisfaction in some types of activities which is as good as any other index available, sometimes much better.

The well-trained diagnostician therefore uses a variety of techniques for gathering data about a person he is going to counsel or concerning whose admission, employment, upgrading, or release he is to make a recommendation He uses psychological tests to obtain information concerning aptitudes for analyzing new situations or for using fine instruments, he checks this evidence against interview material and personnel records which indicate what kinds of new situations the client has met in the past and how he has met them, or what courses he has taken and what hobbies he has engaged in which require manual dexterity and how successful he was in these Ratings and reports from former teachers or employers provide evidence of proficiency in activities not covered by marks and for which no proficiency test data are available They also supply data concerning the ability of the person concerned to get along with superiors, associates, and subordinates, something not assessable by means of the usual psychological tests These illustrations could be extended indefinitely, but should be sufficient to illustrate the point that testing and non-testing techniques need to be used in combination for the effective gathering of psychological and social data

The above discussion presupposes the validity of the psychological tests that are used, just as it presupposes the validity of the other methods of gathering data and of the data which they yield Educators and business men who are not trained in statistics and in experimental methods, and some who are trained in experimentation in other fields but not in psychology, often fail to realize that in a blanket questioning of the validity of tests they assume the validity of some other criterion or predictor such as school marks, supervisors' ratings, production records, or

their own judgment They too often do not know how unreliable or invalid these other indices have been shown to be by objective investigations Ample evidence on this subject will be presented later in this book, in connection with the problem of selecting a criterion and in discussing the validity of each test covered in detail But it is pertinent at this point to introduce some evidence of the value of tests in vocational counseling

The National Institute for Industrial Psychology has conducted a number of studies in England and Scotland over a period of years, in order to ascertain the value of vocational tests in counseling boys and girls in their early teens who were leaving school and taking employment The results have been consistently favorable to counseling which utilizes test data along with other information rather than depending only upon traditional sources of data (11,389,401) Allen and Smith (11), for example, followed up the children who had graduated from four elementary schools A control group had been counseled without benefit of test data, whereas an experimental group had been tested with a variety of vocational tests and counseled in the light of all types of data The vocational adjustment of the experimental group, as evidenced by job stability, satisfaction, earnings, and similar criteria of success, was significantly better than that of the control group

Pitfalls in Diagnostic Testing

Four major types of error are frequently made by users of tests These are 1) the neglect of other methods of diagnosis, 2) overemphasis on diagnosis with the resulting tendency to neglect counseling, 3) failure to take into account the specific validity of the tests used, and, 4) the neglect of other methods of guidance which should normally accompany diagnosis and counseling The first two pitfalls have already been dealt with at some length in this chapter, the third is discussed in the next chapter, in concluding this chapter some remarks on the fourth type of error are in order

Many of the earlier writers on vocational guidance, working at a time when psychological tests were first being developed and when interviewing was an unanalyzed art, were more impressed by the promise of exploratory activities in school and on the job than they were by diagnosis and counseling Aware of the extremely limited usefulness of the tests of their day and of the subjectivity and inadequacy of the interview as then used, they had more faith in the ability of the individual to "find himself" as a result of exposure to a variety of experiences in his school work,

leisure-time activities, summer jobs, and first few years of work than as a result of a counselor's work with him This point of view is expressed as late as 1932 in Brewer's *Education as Guidance* (119), the title of which indicates its philosophy

Not a few more recent writers on vocational guidance have gone to the other extreme, particularly those who have had a part in the development of vocational tests during the past twenty years Impressed by the gains made in our ability to diagnose and predict, they have tended to emphasize the role of the counselor or employment manager and to minimize the importance of exploratory and induction activities This emphasis is shown in the writings of some psychologists of the 1930's (190,928,931).

A third, most recent, group of writers have introduced still another emphasis, that on therapy or counseling at the expense of diagnosis and exploration, the first of which is considered positively harmful while the latter is not considered at all, because of the emphasis on personality adjustment (594,641)

Rogers' and Williamson's points of view have already been discussed in another connection, the point which it is desired to bring out here is that both of these newer emphases have minimized the role of exploration by the individual and the use of exploratory activities by the counselor as a means of furthering vocational adjustment In the opinion of this writer, diagnosis and counseling are essential to a program of vocational guidance, and so is exploration The effective vocational counselor is one who knows when and how to use diagnostic techniques, when and how to rely primarily on counseling, and when and how to help the counselee engage in activities which will help him to obtain the insights and information needed In industrial and business personnel work also, there are circumstances in which good selection is the crucial thing in securing well-adjusted employees, others in which helping them to understand themselves and their situations better is most important, and still others in which good induction into the new company and try-out in a variety of activities are the key to developing effective employees, the most competent personnel man relies on a combination of such procedures To become so absorbed in the mechanics or dynamics of one aspect of vocational guidance or personnel work as to lose sight of the others, or to depend exclusively on one or two rather than using a combination of all three, is to impose an unnecessary limitation upon the effectiveness of one's work

CHAPTER II

TESTING AND PREDICTION IN VOCATIONAL SELECTION

The Peculiarities of Selection Testing

ALTHOUGH the tests used in vocational counseling are often identical with those used in selection, the ways in which the tests are used have generally differed considerably In vocational counseling, the primary objective is the development of an understanding of an individual by himself and incidentally by the counselor, and the relating of personal to occupational data This is by definition a broad task which in our present state of knowledge requires considerable dependence on non-testing techniques and subjectively obtained information concerning both counselee and occupations Perhaps some day the dream of a comprehensive battery of tests and of test weights for all the major occupational fields, described by Clark Hull (385 Ch 14), will be realized, but current opinion is in agreement that both people and occupations are too complex for this to be at all likely In vocational selection, on the other hand, it has proved possible to rely more heavily on testing procedures Familiarity with the reasons for this is essential to the effective use of tests in both counseling and selection

Fundamental among the factors which make possible greater reliance on tests in vocational selection is the relative simplicity of validation, that is, of checking test results against behavior which one is attempting to predict Whereas in counseling one is concerned with a great variety of occupations, in selection the focus is on suitability for one or at most several somewhat related jobs The personnel man interested in improving the selection of employees for certain jobs in his company works with a relatively *uniform criterion group* (men in one job) and with a relatively *simple criterion* He is therefore able to make a careful first-hand analysis of the activities involved in the job, to select or develop tests which seem likely to prove valuable in predicting success in its activities, to check up on the actual value of the tests and of other indices such as

the judgments of interviewers, and to utilize in his selection program the combination of techniques which has actually worked best for the job in question If, for example, the objective is to select effective operatives for a certain type of assembly work, an analysis can be made of the processes involved in the assembly and of the skills which seem to be required by them Possible criteria of successful performance can then be examined, some of them designed to serve as overall indices of success, some perhaps selected to serve as measures of success in special aspects of the work in which specific aptitudes play an important part In an assembly job the overall criterion may be the number of assemblies correctly completed per working day or other unit of time, specific criteria are not likely to be available in as simple a task as assembly work, although some such work can be broken down into processes requiring primarily gross and fine manual skills, spatial judgment, and perceptual speed The frequently forced dependence on one overall criterion of an objective type has the advantage of reducing the amount of experimental work, but has the disadvantage of making it seem deceptively simple Research has shown that production is affected by many factors, including payment methods, location of work, type of supervision, and union policies Despite this fact, the use of vocational tests in selecting employees for one type of job in one company, in which most of these other factors are constant, is made relatively simple by the possibility of one fairly adequate criterion of success.

A third factor which operates to make the use of tests in vocational selection easier and more helpful than in counseling is the fact that the personnel man has some *control over the job situation* As he is working for the company for which he is trying to improve employee selection the company has a stake in his success, and as he knows the situation in which he works, the people whose co-operation he must have, and the policies governing their work, he is likely to be able to obtain the co-operation which he needs and to be able to make changes in policies, schedules, and other aspects of operations in order to achieve his objectives This improves both the chances of developing good tests and the prospects that the personnel whom he has selected will work under conditions which permit the success of qualified employees It should be noted, however, that since the user of tests in personnel selection is part of an operating agency and must fit in with the operating needs of other officials he is subject to pressures which may handicap him in his work Among these are the need for immediate results when preliminary work

should be done before applications are made, the lack of sufficient numbers of employees in some jobs for adequate standardization and validation to be possible, the difficulty of obtaining adequate criteria (e g , the impracticability in some situations of training supervisors to rate objectively), and the fact that certain operations cannot be interfered with in the way necessary to a particular project

The fourth factor which generally operates to make possible greater dependence on tests in personnel selection than in counseling is the practicability and superiority of *custom-built tests* Experience has repeatedly shown that, when a battery of tests is developed especially for use with one job or a group of jobs in the organization, specific local factors can be taken into account which make the tests more valid than tests which have been developed with more varied applicability in mind This is a crucial point which should be borne in mind by every user or potential user of vocational tests for selection purposes, given the time and the highly-trained technical personnel necessary to such work, selection tests developed especially for use with certain jobs in a given organization are likely to prove much more valid than more widely applicable tests A knowledge of the nature and validity of existing tests, such as it is the purpose of this book to provide, is essential to good testing of any kind, but the user of tests for selection purposes needs to master also the techniques of test construction and validation and to apply them to his work, or to obtain the services of a specialist who can, under his general supervision, carry on such work The next chapter contains a discussion of the logic and methods of test construction and validation, but does not attempt to present the statistical procedures As stated in the introduction, that should be the subject of another book

An illustration of the superiority of custom-built tests will help make the point that selection testing is more practicable than guidance testing. In selecting and classifying cadets for training as pilots, navigators, and bombardiers in the Army Air Forces in World War II, some work was done with tests of spatial visualization such as Thurstone's Surface Development (316 273) with results which led to the conclusion that existing tests of this factor were not promising for aircrew selection (rbis= 16 with flying success) Instead, work was begun along lines which were suggested by job analyses and which involved tasks and materials resembling, at least superficially, the tasks in which success was to be predicted One of these tests which factor analysis has shown to be a measure of spatial visualization in a way realistic for aviation (316 479–486) was

entitled the Instrument Comprehension Test In it the examinee read airplane flight instruments such as the artificial horizon and decided which of the presented alternative pictures of an airplane in flight represented the attitude (position relative to the ground) of the plane indicated by the instruments This test had validities of 39 and 48 (two different parts of the test) for the experimental group referred to above

Most clearly a spatial visualization test for aviation, however, was the Visualization of Maneuvers Test (316 277–284) The items in this test consisted of a stem showing the attitude of an airplane and describing the turns, climbs, and dives it next makes, followed by five multiple-choice pictures of the same airplane in varying attitudes. The task was to choose the alternative which indicated how the plane would be flying after completing the maneuvers described This would seem logically to involve the ability to visualize the relationships of objects in space Anecdotal evidence is available in the observation of experienced pilots taking the test and in their comments after taking it, they gesticulate with their hands and sway in their seats as they act out the maneuvers they are attempting to visualize, and say, afterwards, that they "just about twist your hand off trying to do those maneuvers" The correlation of this test with success in flying training has been shown to be 23 (316 283) These results demonstrate considerable validity for single tests, and more than that which characterized the more abstract type of spatial visualization tests

With the advantages deriving from a relatively uniform situation over which he has some control, with a criterion of success which is simple enough to permit validation but broad enough to be related to a number of different tests, and with the greater similarity between test and criterion which results from the ability to use custom-built tests, the personnel man working on selection problems can well depend more on tests than can the counselor who is trying to help people with vocational choices

The Importance of Other Techniques

Although the psychological factors which can be measured in selection are the same as those which can be measured for counseling purposes, there is less reason for thinking that non measurable factors need to be measured in selection than in counseling, and more direct evidence to justify a greater dependence on the factors which can be measured

Numerous studies of the employment interview, summarized by Bing-

ham and Moore (96), have shown that as they normally work there is so little agreement among the judgments of interviewers that employment interviews have little value Since the bulk of these studies were made, improved techniques have been developed which make possible a reasonable degree of agreement between interviewers, these involve training interviewers, standardizing the interview situation, focussing on certain traits or aspects of behavior most readily observable in the interview, and providing standardized scales for the rating of traits or behavior and the notation of substantiating facts Bingham and Moore (96 Ch 2, 1st ed) give an illustration of a form of this last type Despite such improvements experience continues to demonstrate that in many situations interviewing techniques do not contribute much to prediction for specific jobs For example, an aviation psychologist met regularly with a flight surgeon as a member of a board which reviewed the cases of soldiers who made borderline scores on the aviation cadet classification tests This board interviewed these cadets, reviewed relevant material, and decided whether they should be sent on to flying training or disqualified on the basis of low aptitude The board's judgment was proved to be of little value The procedure was soon dropped, and cadets were disqualified on the basis of test scores alone

Another study was made somewhat earlier by the staff of the same Army Air Forces Psychological Research Unit (316 Ch 24), in which a number of clinical techniques, as contrasted with objective tests, were studied in order to determine their validity in predicting success in flying training These techniques included a standardized interview, observation of behavior in an informal "rest period" between tests, observation of behavior in two standardized situations in one of which the cadet took an apparatus test by himself and in the other of which he worked on a spatial assembly test as one of a group of three examinees, ratings of behavior in standard psychomotor tests, and others The correlations between ratings based on these techniques and success in primary flying school were practically zero, except for coefficients of between 15 and 20 for the ratings based on observation in Heathers' Control Confusion Test and on Super's Interaction Test, the two experimental situations designed especially to bring out ratable behavior The interview ratings had no validity, even though made by interviewers who had at least the equivalent of a master's degree in psychology with an emphasis on clinical work The objective tests used in the standard selection and classification battery had validities which ranged from .29

to 51 in the experimental group of 1112 cadets (214 191) Dependence on tests rather than on interviewers' or observers' judgments is clearly justified by these two studies, although it is conceivable that a more valid interview or observation procedure might be devised and personnel trained to use it, as in the work of the Office of Strategic Services (33,558) Finding time for it would then be the problem when large numbers of candidates are involved The AAF program tested cadets at a cost of five dollars per man, whereas the OSS procedure required three and one-half days, a hundred-acre farm, and fifteen professional staff members for a group of eighteen candidates

It should be pointed out that one reason why tests have proved to be more valid than other techniques for gathering and evaluating personal data for the prediction of vocational success is that the tests themselves have been so constructed as to cover material which is often thought of as obtainable only by other methods It is not meant to imply that the tests measured all relevant variables a multiple correlation coefficient of 66 (214 191) makes it quite clear that other factors also were operating in the AAF studies, and the battery of tests avowedly was weak in measures of personality and temperament But the factual material which is normally obtained by means of interviews and questionnaires and then interpreted subjectively was obtained in a Biographical Data Blank (316 Ch 27) devised by Laurance F Shaffer, weighted according to the experimentally ascertained importance of each possible response to each question, and scored to yield a measure of background factors and experiences which play a part in flying success It had a validity of 33 (214 191) The technique was not entirely new it was used in the Civil Aeronautics Administration testing program by E Lowell Kelly (260) and prior to that had become a standard method in the selection of salesmen by a number of life insurance companies In the latter, for example, a positive weight was given to affirmative answers to questions as to whether the examinee was married, had children, or carried insurance, since these were found to characterize men who made good salesmen

Work done in recent years by German military psychologists (245), by Murray and his colleagues at Harvard before American entry into World War II, and by the same investigator and the staff of the Office of Strategic Services during that War (33,558) has demonstrated that there are possibilities in the development of the standardized situation test (see p 529 ff) which should not be neglected in selection programs,

nor, for that matter, in counseling programs The ultimate form of such tests may perhaps not be comparable to the paper and pencil or apparatus tests that we now consider objective, instead, it may combine some of the standardized features of the objective test with some of the subjective features of the interview But, in improving their validity by standardizing the situation and the method of evaluation, psychologists take them out of the category of non-testing techniques and into that of testing techniques A book of this type written ten or twenty years from now may well need to devote a great many pages to the discussion of such standardized life situation tests At present they are experimental and of unknown validity, and so are briefly considered only as a promising technique for the evaluation of personality

The Validity of Selection Tests

The problems and methods of validating tests for selection and counseling purposes are taken up in the next chapter It is pertinent here, however, to examine the evidence concerning the value of tests in the selection of employees, for what has been said on that score while considering the limitation of other techniques has been piecemeal and incomplete

Working with applicants for employment with a utilities company, Wadsworth (905) gave two intelligence tests to an experimental group and no tests to a control group, the former numbering 108 and the latter 594 men and women After employment, by the usual methods in the case of the non-tested applicants, data were gathered concerning their success on the job Employees were classified as outstanding, satisfactory, or problem employees The results, given in Table 1, show the superiority of test-selected personnel in this one enterprise, as only 5 5 percent of the latter were considered problem employees as contrasted with 29 percent of the non-test selected group

TABLE 1

TEST-SELECTED EMPLOYEES IN A UTILITY COMPANY PROVED SATISFACTORY MORE OFTEN THAN OTHERS

Type Employee	*Test-Selected*	*Non-Test-Selected*
Outstanding	33%	22%
Satisfactory	61 5%	49%
Problem	5 5%	29%
Total Number	108	594

Strong used a different type of test with a different type of employment, obtaining his data in the somewhat less satisfactory manner of testing employees already on the job (775 487–498). Despite this his data are impressive, and there is no reason to think that they would have been different if testing had preceded employment Relevant to this topic is his finding that 56 percent of the life insurance salesmen who scored A on his life insurance salesman's scale sold $150,000 worth of insurance per year (enough to yield a living in commissions at that time), whereas only 6 percent of those who made scores of C sold that much insurance

Finally, data from the army aviation testing program of World War II might be cited, because of the unusually large numbers tested, the extensive batteries of tests involved, and the nature of the criteria used Figure 1 shows the percentage of cadets at each ability level (determined by tests) who were eliminated from primary flying training, the first nine weeks of actual flying as a student pilot The trend is obvious at once the short bar at the top shows that only four percent of the 21,474 cadets who entered training between October 1942 and December 1944 with pilot stanines of nine (standard scores expressed on a nine-point scale) were eliminated from primary flying school because of flying deficiency, fear, or their own request, whereas the long bar at the bottom of the graph shows that 77 percent of the 904 cadets who entered training during that same period with pilot stanines of one were eliminated These low-scoring cadets were less numerous than the high-scoring, because of the raising of requirements as the use of tests became more completely accepted and as the progress of the war made smaller quotas of new pilots possible By the end of the war it was possible to accept for pilot training only cadets with pilot stanines of seven This meant that, instead of an elimination rate of 24 percent as in this group of 185,367 in the middle two years of the war, only 10 percent would be eliminated if other factors remained constant

Even more conclusive evidence is available from the experimental group described in Report No 2 of the Aviation Psychology series (214) and by Flanagan (264) As has been previously stated, this group was selected without reference to test scores, the only official requirement being the passing of the physical examination Actually, the group was also somewhat selected according to traditional methods, as they were accepted at a time when the normally enforced standards were well known and the men presumably applied with the thought that they could meet them This is shown by the fact that only 23 percent were not

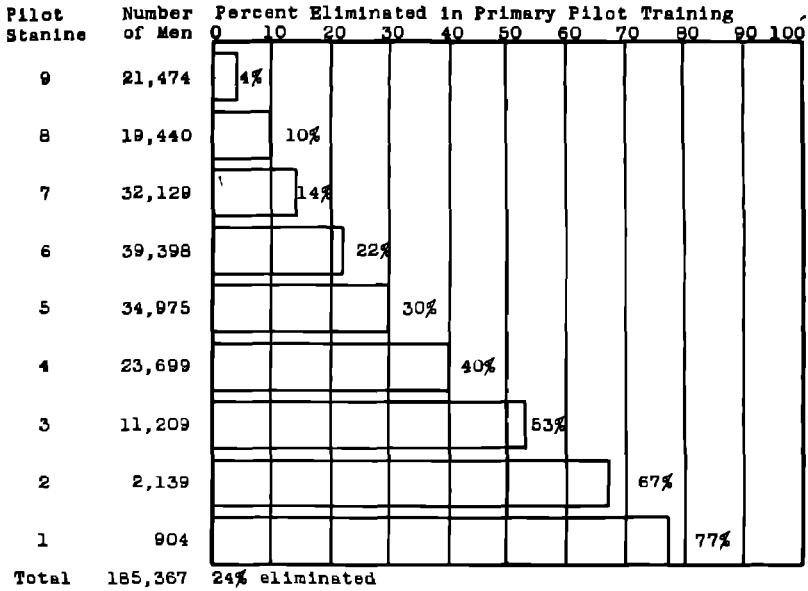


FIGURE 1

TEST SCORES AND SUCCESS IN AAF PRIMARY PILOT TRAINING

The bars indicate the percentage eliminated at each pilot stanine (combined test score), for inability to fly, fear, and at own request Credit for flying experience is included in the stanine Data are for classes trained during 1943 (when some low stanine men were admitted), 1944 and 1945 After Flanagan (264 76)

at least high school graduates, as contrasted with 37 percent of men-in-general at that age (61) Whereas the selection and classification tests normally admitted to training only one failure to every three or four successes, the non-test selected experimental group included one failure for every success If, as there is reason to believe, other things such as the strictness of instructors, check riders, and elimination boards remained relatively constant, the use of tests was clearly an improvement over selecting merely on the basis of physical examination and, to a lesser extent, education

Programs of Testing for Selection, Placement, and Upgrading

Despite the evidence which shows that subjective methods of evaluating applicants for employment add little or nothing to the predictive value of well-constructed and validated objective tests, personnel men

and vocational psychologists continue to utilize interviews, application blanks, rating scales and letters of recommendation in selecting employees This is partly because of an unreasoning distrust of purely objective methods, partly because of the knowledge that even the best of test batteries do not cover everything and the hope that other methods will supplement them, and also because, in practice, tests are often used without the thoroughgoing standardization and validation procedure which is necessary before one can know just how valid they are and whether or not selection is in fact improved by supplementing tests with other techniques

When job analyses have been made the emphasis in testing is likely to be on *placement* on the right type of job, when differential ability data are lacking, it is likely to be on *selection* of generally promising employees

One large corporation, to cite a concrete instance, uses psychological tests in three of its divisions In one division of this corporation a new plant had been built and the personnel director was told that the management wanted to make it a model plant He was accordingly directed to devise a battery of tests which would be appropriate to the jobs to be filled, and to select employees on the basis of tests and other data from the beginning As is frequently the case in actual operations, the pressure of the situation, that is, the need for selecting employees on *some* basis and the belief that even tests which had not been validated in that plant would help in the selection of better employees than would be selected without test results, caused the use of tests without the benefit of the scientific preliminaries which are usually considered desirable The personnel director therefore put into use a battery of tests which, judging by results in other plants in which somewhat similar work was done, seemed likely to prove valuable They were used in an attempt to exclude from any type of job the most awkward, most maladjusted, and least intelligent, that is, for *selection* At the same time provisions were made for the gathering of data concerning the success of the new employees on their jobs Although the use of the tests in making decisions concerning selection could be expected to reduce the range of abilities in any one job, it was felt that the shortage of labor would result in a spread of abilities sufficient to reveal whether or not a relationship existed between test scores and job success In such a situation it was only natural that previous experience, schooling, and similar background factors were weighted

quite heavily by the employment manager in going over the results of tests, interviews, application blanks, and letters of recommendation.

In another division the psychologist in charge of testing began by making a systematic analysis of the jobs in question, using standard job psychographic techniques He then selected and devised tests which he thought would be effective measures of the characteristics which appeared to differentiate the major types of jobs The experimental battery of tests was administered to all applicants for factory employment and, as data accumulated, the test results were correlated with supervisors' ratings in order to determine their actual value in selection One test was found to add nothing to the predictive value of the battery, but, as it took little testing time and appealed to applicants and foremen, it was retained, other tests which had some value were weighted accordingly and used in selection and in *placement* in appropriate jobs The validities of the battery average about 50 and, at the time of writing, are based on rather small groups No personal history or biographical data form of the type discussed earlier in this chapter is used The testing program is still relatively new in this plant For these reasons the employment interview is depended upon rather heavily, and decisions are made after the background and manner of the applicant have been mentally (rather than statistically) weighted in combination with the test results by the employment manager, with the emphasis on placement in a suitable job

The third division of this corporation operated in a part of the country in which the labor shortages resulting from wartime and postwar developments were serious In practice, employee selection became more a matter of employee *placement* The personnel manager therefore selected a battery of tests without regard to special aptitudes and abilities such as might be important in selecting for or in placing people in different types of jobs, believing that even selective placement was generally out of the question in that plant, the emphasis was placed on tests of certain basic general factors the understanding of which would help foremen and supervisors to induct and handle the new employee more effectively The employment battery therefore consisted of a test of general intelligence, a measure of personality adjustment, and a measure of vocational interests The nature of the tests was explained to supervisors, the scores of each new employee were discussed with them, and they were helped to understand the types of adjustment problems which the new employee might encounter It was believed that the supervisors' interest

in and intelligent use of this information was an important factor in the development of satisfactory employees, although no objective evidence was gathered on this subject

Psychological tests are frequently put to use in business and industrial personnel work for the *upgrading* of personnel, that is, the evaluation of employees for possible promotion to more responsible positions In this type of work two approaches are possible, one of them comparable to selection testing, the other to placement testing In the former, tests and other techniques are used which will throw light on the general promise of the persons in question their general intelligence, personality adjustment, leadership, and similar general characteristics are assessed by means of tests, inventories, ratings by superiors, and interviews In the latter, data are gathered by similar methods, but they are data about special abilities, interests, and personality traits that are known or thought to be important to success in specific jobs at higher levels

For example, a number of aviation psychologists worked under the leadership of John C Flanagan in the American Institute for Research, on the evaluation of airline first officers for possible promotion to captaincies In this program an analysis was made of the abilities and characteristics needed by the captain of a commercial airliner Tests were selected which previous work with pilots had demonstrated to be correlated with success in flying twin and four engine planes, others were constructed to measure characteristics not covered by existing tests, and interview procedures were developed for tapping other factors which could most effectively be assessed in face-to-face contacts Techniques for quantifying the results of interviews were developed, and the results obtained by any one interviewer were so treated as to make them comparable to the results obtained by others, thereby minimizing the subjective elements At the same time, the flight records and ratings of first officers by captains and check pilots were utilized as objective measures of proficiency and achievement, after they had been subjected to a statistical study which demonstrated their reliability and validity The resulting data were weighted to provide an overall score indicative of the pilot's promise as a captain, this, and a three hundred word sketch verbally summarizing the first officer's assets and liabilities and pointing out how they might be respectively utilized and corrected in this and other possible jobs, were turned over to company personnel officers for use in making decisions

In such a program tests play an important part in assessing character-

istics which are not called for in the job currently held, or the exercise of which cannot be well observed on the job They help to isolate factors which, even though observable in the employee at work, are so intertwined with other factors that the observer has difficulty in determining the relative importance of a given strength or weakness And, finally, they are free from the taint of possible bias

CHAPTER III

METHODS OF TEST CONSTRUCTION, STANDARDIZATION, AND VALIDATION

TO BE fully competent in the use of vocational tests it is necessary to know all stages and types of work with tests This does not mean that the vocational counselor or personnel director must be an expert in test construction, nor that the developer of tests must also be expert in using them in counseling or selection But it does mean that the vocational counselor must be familiar with the procedures and problems of test construction, and that the technician whose function it is to develop tests must understand their use in counseling and selection, if the tools essential to diagnosis are to be worth using and well used It is therefore the purpose of this chapter, not to provide a manual of test construction, but rather an orientation to test construction which will enable the user of tests in counseling and personnel evaluation to read the published test research with a critical appreciation of the problems involved and thus to understand more completely the meaning of the results obtained when using tests

The development of a vocational test can be broken down into seven major steps These are job analysis, selection of traits to test, selection of criteria of success, item construction, standardization, validation, and cross-validation. In any given test construction project one or more of these steps may conceivably be slighted or omitted altogether when this is the case, however, it should be because sufficient work has already been done along those lines to provide a basis for the next step, or because the pressure of time and circumstances makes the taking of short cuts necessary and dependence on hunches seem wise The critical reader must judge for himself whether or not the omission of the steps was justifiable and whether or not the data are usable The seven steps will now be taken up in some detail

Job Analysis

Before tests can be selected or constructed for the measurement of aptitude or personality traits which affect success or satisfaction, it is necessary to have an understanding of the characteristics and abilities which play a part in the work in question The process of collecting and analyzing information which provides this understanding is called job analysis Whether it is done scientifically or otherwise, some type of job analysis has to be performed before an aptitude test can be constructed It may be an armchair analysis, in which the test constructor draws on his familiarity with the job or occupation for which tests are being constructed in order to set up hypotheses as to the characteristics which make for success in that work. It may involve bibliographical research, (1) to ascertain what others have thought or found to be important in that occupation It (2) may be an analysis of manuals used in the training of people for the work in question, in order to judge the abilities needed in mastering the fundamental skills It (3) may involve discussing it with supervisors, observing and interviewing workers doing the work, trying the operations oneself, or even (4) learning the job and working at it for a period

In analyzing the work of military pilots a combination of these methods was used as time and circumstances permitted First, J C Flanagan analyzed the proceedings of boards which eliminated failing aviation cadets from primary flying training, in order to ascertain the reasons given for their failure by the boards This resulted in a list of characteristics ranging from lack of co-ordination to poor motivation, and a table showing the incidence of each of these reasons in a large sample of eliminees Then J K Hemphill, drawing on his own experience as a civilian flyer, and the writer, depending on observations of military pilots at work and demonstrations of flying in which he performed some of the operations, made an analysis of training manuals in order to describe the pilot's tasks as a basis for setting up hypotheses concerning characteristics which would make for success in learning to fly After this, N E Miller, J L Wallen, and the writer went to a military flying school in which Miller and Super worked as participant observers, living in barracks with the cadets, attending ground school and physical training, handling planes on the flight line, learning to fly, and being graded for their flying on the same basis as cadets Wallen worked in the station hospital, administering clinical tests to the cadets being studied, inter-

viewing them concerning their background and development, and collecting other types of information from hospital, training, disciplinary, and other records. All three job analysts kept notes concerning the observed behavior of the twenty cadets of whom intensive case studies were being made, whether on the flight line, in the barracks, or on "open post" in the nearby town The two investigators who were flying kept detailed records of their own experiences in learning how to fly These materials provided a basis for detailed study of the task of learning to fly, of emotional aspects of the experience of learning to fly, and of factors which made learning to fly easier or more difficult for a random sample of cadets P L Fitts interviewed the returned members of a bombardment squadron in order to get their account of the nature and requirements of combat flying, analyzed the material, and made it available to aviation psychologists working on test construction Flanagan spent some time in a combat theater studying records, interviewing flyers, and flying a number of missions in order to analyze the task of combat flying at first hand Later, research detachments conducted similar investigations on a larger scale in most theaters of the war (467)

The above description of job analysis activities in one practical situation is given in order to illustrate the variety of approaches that may be used in the study of the nature and requirements of a job or an occupation In practice there is not necessarily one method of job analysis, it is more likely that there are several which will yield valuable information, and that more than one must be used if adequate data are to be made available as a basis for selecting or devising tests The brief survey of the development of job analysis methods which follows will bear this out

The scientific analysis of jobs was begun early in this century by Frederick W Taylor (811) as a means of increasing the productivity and facilitating the work of industrial employees It was soon seized upon by psychologists as a method of ascertaining in a preliminary way the abilities and traits needed in an occupation and thus of providing a basis for test construction Taylor's methods, and those of Gilbreth (289) and other workers whose interest was primarily in engineering, emphasized time and motion study, the picture of a job derived from such work therefore proved to be too narrow in its viewpoint for personnel work, leaving out of consideration such things as the education and training required of the worker, the interests which might find outlet in the activity, and the environment in which the work is done They also provided too detailed a picture of the manual operations involved in the work, al-

though Cohen and Strauss (162) have used the technique effectively in studying manual dexterity. Other methods were therefore resorted to, in an attempt to obtain information which would provide a suitable basis for test construction

One of these was the job psychographic method developed by Viteles (581,899,901) It begins with a description of the occupation, describing the duties performed, the nature and conditions of the work, the training involved, the related jobs from which workers may be recruited and to which they may be promoted, the advantages and disadvantages of the work, and the personal, physical, educational, temperamental, and experience requirements of the job This material is gathered by observing the performance of the work and by interviewing workmen and supervisors So far, this is the standard job description or position description technique In order to objectify the analysis of the job Viteles developed a standard list of 32 abilities which are rated on a five-point scale by the analyst, the list consists of such factors as energy, co-ordination, visual discrimination, and logical analysis The ratings, placed on a graphic scale, yield a profile of the abilities required by a job and give their name to the method

The most recent form of job analysis, adapted especially to vocational guidance because it deals with broadly rather than with narrowly defined jobs, is that widely applied by the Occupational Analysis Division of the United States Employment Service under Carrol L Shartle (714 Ch 11) Items which have a bearing on test construction include a description of the work performed, the amount and type of supervision received, the responsibility, knowledge, initiative, alertness, judgment, dexterity, and accuracy involved, the tools used, production standards, working conditions, physical demands, and other characteristics required for performance of the work The use of this procedure, like Viteles', yields a list of abilities and traits which are considered important in the occupation or job being studied

Selection of Traits to Be Tested

The analysis of the job provides the test constructor with a list of aptitudes and traits which are deemed important in that job. But this list is subject to two serious limitations These are the subjectivity of the evidence and the uncertainty that a particular factor, even if it proves to be important, will differentiate this job from others The fact that ability to get along with others is thought important in a given job is, for ex-

ample, ascertained only by the analyst's observations or by opinions transmitted to him by persons who know the work The data are no more reliable than the judgment of the people gathering or supplying them Furthermore, if the presence of the trait is subjectively ascertained in the first place, there may be no objective method of assessing it, for it may be a characteristic which has so far eluded the attempts at measurement Granting that ability to get along with others is a prerequisite of the job being studied, there is still a question as to whether or not it differentiates this job from others There are many jobs which require ability to get along with others, even if this trait could be measured, its measurement might contribute little that is of value to differential diagnosis and prediction

Once the job analysis is complete and the list of presumably important characteristics is available, the first task of the test constructor is to make some decision as to, 1) the relative importance of each trait or aptitude, 2) the availability of a suitable criterion against which to validate a test of this trait, 3) the chances that a given trait is important in this job and unimportant in others with which he is also concerned, 4) the unavailability of some reliable and economical non-testing technique for judging this characteristic and, 5) the prospects of his being able to locate or devise a test which provides an objective measure of the characteristic in question The job analysis should provide evidence of a subjective type concerning the first point, as, for example, in Viteles' psychographs The next section deals with the important problems which arise in connection with the choice of criteria A comparison of the job analysis data for the job in question with available evidence from other jobs should provide a basis for judgment of the third point In connection with the fourth point, the use of school grades and supervisors' ratings should be considered For the fifth, the psychologist must be well acquainted with the various types of tests which are already in existence and with the extensive literature on test construction in which abortive as well as successful efforts at test construction have been described In the light of these considerations, the psychologist is able to draw up a list of aptitudes, skills, and personality traits ranked in the order of the likelihood with which they may be successfully studied

Selection of the Criteria of Success

Jenkins (400) has pointed out that the events of World War I taught American psychologists the *necessity* of validation, the next two decades

taught them much about the *technique* of validation, and that World War II drove home the necessity of devoting much time and thought to the *basis* of validation In most of the test validity research of the 1920's and 1930's much space is given to descriptions of the technique of test construction, the methods of securing data, the description of the criterion used, and the results of the relating of test scores to criterion data. Not infrequently one of these topics is somewhat neglected—that in which the criterion is described But, even when the criterion is adequately *described,* too little attention is paid to its adequacy as an index of success

This lack of emphasis on the criterion can be illustrated by a study (669) in which the group of aircraft factory inspectors on which the battery of tests was validated were not defined as to type of material inspected, sex, or age, and were described as "probably representative" with no supporting statistical analysis, the raters who made the criterion judgments knew the subjects as students in a refresher course, but knew their job performance only in "most" cases, the ratings of two instructors had an intercorrelation of .77, and their correlation with subsequent ratings by supervisors was .42 Some of the data just presented are quite adequate, the intercorrelations of ratings being quite high for such material, and yet it should be obvious that, with no more attention devoted to the criterion than in this study, it is difficult to interpret the results For example specifically what type of performance was rated, that it correlated highly with intelligence (.64) and only moderately (.32) with mechanical comprehension? Data for engine and fuselage inspection might differ Was the immediate criterion (instructors' ratings) only moderately related to the ultimate criterion (supervisors' ratings) because of low reliability of the latter, lack of common factors in the instructional and work situations, or some other uninvestigated factor? Admittedly the judgments of the instructors are one type of evidence that is available early in the new employees's job experience, but how valid a *criterion* is it, that is, how good a measure is it of what the tests are trying to predict? If a test has a correlation of .64 with the immediate criterion, and the immediate criterion has a relationship of only .42 with the ultimate criterion, the relationship between the first predictor and the ultimate criterion is not very high A more thorough study of the nature and meaning of the criterion serves to clarify issues and suggest better predictive devices At the same time, it is true that whether or not it is desirable to devote time and personnel to such a study depends on other

factors in the situation, e g , the savings that would be effected by improved procedures

The typical but unwise procedure in test construction is, too often, to leave the detailed consideration of a criterion until somewhat later in the process than has been done in this discussion Usually, having decided what factors he should try to test, the psychologist has proceeded to develop suitable tests, administer them to appropriate subjects, and then for the first time seriously consider the problem of criteria The vague ideas that he has so far had are now crystallized, the most readily available index of success is used with little or no investigation beyond a cursory check on its reliability, and the relationship is computed

The experience of Naval aviation psychologists summarized by Jenkins in the paper referred to above, and the experience of Army aviation psychologists summarized by R L Thorndike (833), suggest that the order of the steps taken in test construction needs to be changed, and that considerable emphasis needs to be put on the problem of selecting and evaluating a criterion early in the process Once the traits to be measured have been determined, attention should be turned to the selection of a criterion and to the refinement of methods of collecting evidence against which the tests to be developed can be validated The discussion which follows describes the major types of information which are used as indices of vocational success, indicates some of their strengths and weaknesses, and illustrates them from research In doing so, it relies to a considerable extent upon the work of aviation psychologists in World War II, partly because the nature of the aviation psychology programs, both as to problems faced and staff available to study them, makes them an especially good source of such material Illustrations are also taken from studies in the field of industry and education

Thorndike (833 Ch 4), Humm (386) and others have distinguished between *immediate, intermediate,* and *ultimate* criteria In military aviation these are respectively illustrated by such evidence as ability to complete training as a bombardier, accuracy of bombing (indicated by average circular error) on the practice range in operational training, and accuracy of bombing in combat Immediate criteria are generally partial, that is, they tend to emphasize limited aspects of performance If grades in medical school, for example, are used as an index of success, some men with good academic ability but poor social adjustment will be rated as more successful than certain other students with somewhat less academic ability but superior social adjustment, whereas if an ultimate criterion

of success in the practice of medicine can be utilized the latter may prove to be more successful than the former Conversely, ultimate criteria are more complex than immediate or intermediate indices of success, for this reason, as well as because of the pressure of time, they are rarely used in test validation In the case of military pilots, for example, it was necessary to put a classification program into operation on a large scale shortly after the bombing of Pearl Harbor This meant that there was no time in which to gather data on the subsequent combat success of cadets before establishing weights for the experimental tests Collecting such data actually took more than two years Instead it was necessary to use an immediate criterion, in this case evidence of the cadet's ability to graduate from primary flying school, which became available in about five months This is by no means a simple criterion, as it is affected by a variety of factors such as the cadet's various abilities and personality traits, the attitudes of the instructors under whom he works, and the extent to which the school he attends adheres to or deviates from established practices and standards But it is not an ultimate criterion, as ability to complete the first stage of flying training is not necessarily identical with ability to outfly enemy pilots or to withstand the greater and more enduring stresses of battle Since pilots who cannot complete training never get to combat the criterion is, however, suitable in a negative way The same argument applies to the selection or guidance of physicians, teachers, and any other group which must surmount a training hurdle before they can compete in practice

The first characteristic to be sought in selecting a criterion is *relevance* If the immediate criterion is to be a valid one, it must adequately represent important aspects of the ultimate criterion If success in completing training is to be a suitable immediate criterion, the activities and requirements of the training program must resemble those of the job Fortunately, the job analysis should provide a fairly good basis for a subjective judgment of this matter Jenkins (400) cites the case of aerial gunnery, in which intelligence test scores were found to correlate highly with grades in training, and might therefore have been assumed to predict success in actual combat, but when the curriculum was revised to make it less abstract and more practical the correlation between intelligence and grades fell to zero

A second characteristic of a good criterion is *reliability* (see page 651 for definition) Thorndike (833 34) has pointed out that although high reliability is not essential in a criterion, provided it is stable enough to

reveal the existence of a relationship, the more reliable the criterion is the more clearly the degree of the relationship is demonstrated Low reliability is caused by intrinsic factors such as the inconsistency of the performance which is being studied, and by extrinsic factors such as variability in the conditions of work, the lack of agreement between raters either in the use of terms or in the interpretation of behavior, and bias in the situation An illustration of *inconsistent performance* is provided by an analysis of errors in determining the position of an airplane at key points in the mission (833 44), which showed that the number of such errors made in one mission has no relationship to the number of errors made in the next mission As the reliability of performance on a single mission was considerably higher, it is probable that both the inconsistency of the performance of such a complex task and variations in external conditions played a part in the unreliability of performance from one mission to the next *Variability in the conditions of work*, in these same aviation studies, consisted of such factors as temperature, visibility of targets, and turbulence of the air and consequent instability of the navigator's and bombardier's working platform In business and industrial studies such variations are illustrated by differences between selling on an open floor on which the customer can approach the merchandise and the clerk can use his skill in approaching the customer, and selling behind a counter where the clerk can merely await the customer in a more passive way, or by differences in supervision which affect the attitudes and output of the workers Meltzer (524) has for example reported a study in which the Minnesota Rate of Manipulation Test (Placing) had a correlation of — 27 with output under one management, and of more than 20 in the same department under a different type of management and with the different attitudes which it engendered The *lack of agreement between raters* is so well-known a factor that it hardly needs elaboration Jenkins (400) mentions a study in which Naval aviation cadets were given successive check flights by two experienced instructors, with a correlation coefficient of approximately zero for the two sets of grades *Bias in the situation* is well illustrated by differing standards in the judgment of performance in different training institutions from which graduation is the index of success, for example in traditional academic colleges on the one hand and in progressive colleges which emphasize more than intellectual accomplishment on the other.

Criteria may be classified as *proficiency measures, output records, ratings, self-ratings, administrative acts,* and *internal consistency meas-*

ures As Thorndike points out in his volume of the aviation psychology series (833), some of these are enduring records which can be scored with perfect agreement by different workers at different times (the first two categories), such as answers to a multiple-choice test or hits on a target, some leave no enduring record but can be recorded objectively by an observer (administrative acts, ratings and anecdotes), such as number of bounces in landing a plane or number of customers approached, and some are subjective evaluations for which no objective evidence of any type is available save the overall impression in the observer's mind (ratings) Some discussion of each of these categories, with illustrations of their use, should provide a better understanding of the validity of tests

Proficiency as measured by tests of information and skill in the performance of a task is sometimes used as an index of success In some occupations, the work of which closely resembles the work of the proficiency test, this type of criterion may be quite appropriate The work of a navigator in flight resembles that of the student of navigation in the classroom in many important respects, even though it may differ insofar as working conditions are concerned The computations and instruments, and even the sequence in which they are used, can be made the same in the classroom or group test as in the airplane This logical analysis is borne out by a correlation of 49 between final examinations in ground school and final average grade for missions (265 122), although the coefficient is low enough to make it clear that there are factors operating in flight which do not operate in the classroom, probably factors of an emotional and perceptual nature In many other occupations the proficiency test situation is too unlike that in which the actual work is performed for it to seem a satisfactory criterion knowledge of the operation of a 50 caliber machine gun, for example, would not appear to involve the same aptitudes and skills as ability to hit a moving target with it while standing on an unstable moving platform Before an achievement test can be considered a good criterion of success, an analysis of the job and of the factors covered by the test is necessary.

Output can be gauged in a number of ways, varying with the nature of the task, In a production job it may be the number of units produced per hour, whether the units are identical parts turned on a lathe or pounds of butter wrapped, or it may be the average earnings over a given period when wages are based at least in part on volume produced In a sales job it may be the number of units sold or the dollar value of

the total sales, or a ratio of sales income to sales expense In military aviation it may be the number of hits on a target in gunnery, the average circular error in bombing, or the number of planes shot down by fighter pilots or gunners Criteria such as these seem delightfully concrete and objective at first glance, but one of the bitter lessons learned by applied psychologists engaged in test construction work is that the appearance of objectivity is frequently deceptive

Investigations of incentive systems have shown (514,637), for example, that the output of industrial workers is often governed by factors other than individual differences in abilities or motivation and that artificial limits are often set upon the amount produced per worker per hour. A detailed study by Rothe (653,654) showed that individual daily work curves of butter-wrappers vary greatly, but that nevertheless group trend lines were a stable and usable criterion He found no evidence of restriction of output in his subjects In sales work differences in territories, in type of clientele, and in the aspirations and circumstances of the salesmen often attenuate the relationship between volume of sales and abilities Strong (772) investigated the case of a life insurance salesman whose annual sales were not as great as would have been anticipated of one with a test score as high as his It developed that he had a private income and therefore aspired to sell only enough insurance to supplement his income In executive jobs company policies greatly affect the amount earned E L Thorndike (831 86) reports the cases of two presidents of equally important and well-known corporations, one of whom received a salary of $420,000 per annum, the other $125,000

While making a job analysis of flying it occurred to the writer and a colleague that a pilot's ability to hit a target in air-to-ground and in air-to-air firing should be a good index of flying skill, as the fixed gunnery engaged in by a fighter pilot involves pointing the airplane and maintaining it as a steady platform while squeezing the trigger It would, it was thought, have the unique advantage of being an entirely objective index of flying skill obtainable before combat It had the further advantage that gun-camera photographs could be used, further simplifying and objectifying the scoring After some preliminary studies a large scale study made under Neal E Miller's supervision at Randolph Field showed that the reliability of air-to-air gunnery scores was 63 when 1200 rounds were fired, and that air-to-ground scores had a reliability of 59 when based on 400 rounds (833 52) While these reliabilities are high enough for use in validation studies, they are surprisingly low for something as

objective as ability to hit a target, and they are among the best of such results A study of the reliability of bombing scores, also cited by Thorndike (833), reports a median reliability of 08 As Kemp and others have shown in the original studies (421 42–52), so many factors enter into the accuracy with which bombs are dropped that one cannot predict the performance of a given bombardier from one mission to the next unless he flies with the same crew and the personnel factors are thereby kept constant, even then, weather provides a vitally important but extraneous variable

Output may also be judged somewhat more subjectively, by having experts evaluate the product as to quality This is done by developing a score sheet on which specific aspects of the work are rated and the total score obtained by combining these ratings This is a method commonly used in evaluating school systems and in phase checks or performance tests for aerial gunners, but it has not often been applied to civilian jobs The work to be evaluated need not be tangible, but may instead be simply an observed performance as in the case of the standard flight checks developed for pilots in the Army Air Forces In these flight checks the cadet performs certain highly standardized maneuvers, while the check pilot or examiner records such objectively determined items as the angle of bank in a steep turn, the time taken to complete it, and changes in altitude These observed performances provide an objective basis for the performance score Work along these lines did not progress far enough for complete evaluation before the end of the war but one group of 16 selected items had a reliability of 39 for cadets with 15 hours of training and 50 for men with 55 hours of flying (833 47)

Ratings of performance provide a widely used type of criterion, probably the most common because of the relative ease of obtaining them The history of ratings has, however, been extremely disappointing, and when they are relied upon today it should be only because of inability to find or devise a better criterion and after systematic steps have been taken to make them as reliable as possible The literature on rating as a technique is too well known to need reviewing here, it is well treated in Symonds (810 Ch 3), Strang (768 Ch 6), and Traxler (860 Ch 7) The recent work on the California Adolescent Growth Study (567), although not concerned with vocations, provides suggestions for further improving rating scales and their use From the point of view of the reader of the literature on the validity of tests, the questions to be kept in mind have to do with the extent to which the ratings of one judge

agree with those of another, the possible influence of halo effect (the tendency to rate specific traits on the basis of an overall evaluation), and the relevance of the traits or behavior actually being measured to the work in question In one study (833 50–51) in which airplane commanders were rated while going through operational (combat) training, the rating for "likeableness" had the highest correlation of any of the ten traits rated with the overall rating of suitability for combat flying There would seem to be little relevance in this case, and considerable halo effect

In studies of the use of tests in vocational counseling conducted in England under the auspices of the National Institute for Industrial Psychology (11,232,389,401), and in a few American investigations (164,706) ratings of vocational adjustment have been used as a criterion In these instances the investigator usually makes a case study of the individual in his work and gives him a rating for vocational adjustment according to the extent to which he seems to be properly placed, satisfied with his work, and satisfactory to his employer Little attention has as yet been paid to the adequacy of the judgments made by such investigators, presumably because of the labor involved in having more than one judge go over the necessary case material In many respects, however, this would appear to be an ultimate criterion of so desirable a type as to justify giving time to devising more economical ways of using it and more thorough study of its reliability

Most users of ratings have obtained ratings of the traits or behavior of individuals In a few investigations the focus has been not on a person, but on some tangible product of that person's work When this has been the case the results are somewhat more encouraging One of the best examples is the Minnesota Mechanical Abilities Project (588 201), in which industrial arts teachers rated the shop products of junior high school boys for quality of workmanship In such rating the identity of the worker can be disguised to avoid halo effect, thereby focussing attention on the specific aspects of craftsmanship to be judged The reliability of the ratings in this study was 76 in the woodshop and 72 in the sheet-metal shop The principal weakness in such criteria, as in the case of more objective output criteria, is the neglect of important human factors not directly revealed in the product of the worker.

Self-ratings have occasionally been used as a criterion of success in attempts to get at the less tangible and more personal aspects of vocational adjustment (377,667,790) The focus in these investigations has generally been on the nature and extent of job satisfaction rather than on the

predictive value of tests, although Sarbin and Anderson (667) did study the relationship between Strong's Vocational Interest Blank and satisfaction in work In studying the value of tests in vocational selection, the emphasis is appropriately on the effectiveness of the worker in performing his task as indicated by ratings of supervisors or by output, but as the use and study of vocational tests in counseling is improved it is probable that more attention will be paid to ratings based on case studies and to self-ratings, the former as an index of overall vocational adjustment, and the latter as a criterion of the worker's feelings of success and satisfaction in his work As self-ratings of job satisfaction such as are provided by Hoppock's scale and the occupational adjustment key of the Bell Adjustment Inventory are further refined, to distinguish between *job* and *occupational satisfaction* and between the various components of each of these global concepts, they will probably find increasing use in the validation of tests and inventories for vocational guidance

Administrative acts which provide criteria of vocational success include the obtaining of employment in a given field, promotion, increase in pay, discharge or failure, and other tangible evidence that people employed in the field consider the individual in question a success or failure These administrative acts have many of the drawbacks of ratings, and are in fact administrative outcomes of ratings, on the other hand, they are generally made after more serious deliberation than a rating is, because of the obviousness and immediacy of their effects on employer as well as on employee Ability to complete flying training was thus the best immediate criterion of success in the Aviation Psychology Program of the Army Air Forces, promotions, decorations, assignment to first or co-pilot duties, assignment to lead crews, removal from flying status for flying errors, and removal from combat because of operational fatigue (neurotic reactions to combat stress), were also used as intermediate and ultimate criteria (833 55) The National Institute for Industrial Psychology has frequently used ability to keep a job as a criterion (11,232) in a period of depression, when jobs are scarce and promotions come slowly, this is presumably a sound criterion, but in more prosperous times, when transfers to better jobs are more easily obtained, and when the scarcity of labor makes employers retain marginal and submarginal employees, the criterion is obviously less adequate This illustrates the defect inherent in all administrative criteria, that is, the degree to which they are affected by external factors Ability to complete a training sequence may depend in part upon changes in standards from one time to

another or from school to school at one time, for example, one primary flying school consistently eliminated 50 percent of its students and another only 10 percent, despite control of the quality of the cadets sent to them for experimental purposes without their knowledge (316 116) In the last analysis, administrative acts make a good criterion because it is in terms of them that success and failure are judged in daily life, at the same time, it is important for the user of tests based on such indices to know just what factors were operating in the administrative situation at the time in question, and the effect of their presence on the criterion and on the test validities

Internal consistency (see page 652 for definition) is frequently used as an index of the validity of a test, although it has no necessary significance for vocational prediction In the case of general intelligence, the vocational significance of which has been demonstrated in numerous studies with a variety of tests and for the measurement of which certain types of items have amply been demonstrated to be effective, it may be sufficient to check the internal consistency of a new test and to standardize it on a good sample population for its results to be useful in vocational guidance Ascertaining its validity for specific occupations would be helpful to counselors, but might be dispensed with if it interfered with better validation of other tests On the other hand, measures of special aptitudes, of interest, and of personality are still so little understood, and the nature and operation of these charcteristics in determining vocational success and satisfaction is so uncertain, that merely knowing that the items in a test measure the same thing is insufficient The score on a test should be a measure of one characteristic rather than of several unrelated traits or abilities, and the people who score high on one half of the test should score high on the other half, in order that one may be sure the test is measuring something and measuring it well, but the vocational counselor, psychologist, and personnel man need to know that what is being measured is related to success in the activity or activities in question This requires an external criterion of validity such as those discussed in the earlier paragraphs of this section

Knowing the various types of criteria discussed above, and their advantages and limitations, the test constructor canvasses the situation in which he is working to ascertain what kinds of criteria are already available to him, and which could be made available if proper steps were taken Existing criterion data are analyzed in order to ascertain their reliability Supervisors who already rate their employees may be given

a refresher course in rating in order to make their results more reliable, or statistical corrections may be made for constant biases which the data have revealed in certain raters Production records may be usable in their present form, or it may be found that there is too little variation among workers for them to serve as success criteria If no suitable criterion already exists, the psychologist must decide which possible criterion lends itself most effectively to use in that situation and how data might be collected He may need to use a second-best criterion, because the data are more readily gathered than those needed for the best possible index In any case, it is important that the criterion chosen be not only obtainable and reliable, but also appropriate to the test or tests being validated relevance should not be sacrificed to convenience or to objectivity These decisions tentatively made, the next step is the building of apparatus or the writing of items

Test Construction

Once the nature of the characteristic to be tested and of the criterion to be used in validating the test have been decided upon, the *choice of type of test* and of test item is relatively easy If the characteristic to be tested has been isolated by job analysis procedures it may be a relatively complex bit of behavior requiring a miniature situation test and therefore, as a rule, apparatus Or the characteristic may have been broken down into relatively abstract components which lend themselves to paper and pencil testing thus in aviation cadet testing a large fraction of the validity of certain apparatus tests lay in their measurement of spatial visualization, a factor which was well tested by paper and pencil tests used in the same battery (315,316,319) Knowledge of the literature of aptitude and personality testing is also a source of ideas as to how to attempt to measure a given trait

The type of test having been decided upon, the next step is to *construct* the apparatus or to *draw* or *write* items In the case of an apparatus test first a sketch and then a rough pilot model is made in order to devise suitable mechanical or electrical methods, to ascertain the most effective size or sizes for the various parts, and to have a model for use in experimental trials In paper and pencil test construction the procedure is to draw up an outline of the proposed contents of the test or inventory, write, photograph, or draw items of those types, and refine them by checking and rechecking Thus in constructing a three-dimensional test of spatial relations one would cut blocks of wood of various sizes with various

degrees of complexity, in order to ascertain which yield the best results; in the case of a general information test one canvasses encyclopedias, current magazines and newspapers in order to choose topics for items, and makes up questions with suitable right and wrong answers

The preliminary form or forms of the test having been prepared, the test is *tried out* on a small group of subjects, who may be a sophisticated group of co-workers or a sample of the type of subjects for whom the test is designed Ideally, both are done in order to get subjective comments and criticisms from the points of view of both test constructors and persons like those to be tested In one project, for example, the author helped devise a personality inventory for aviation cadets The topics covered had been selected by the test construction staff, items had been suggested by cadets in free response answers to somewhat general questions about their satisfactions and complaints in the Army, and questions had been framed and multiple-choice answers put in tentative form by the test constructors This preliminary form was then administered to a small group of aviation psychologists who had not worked on it and to several small groups of cadets, who were asked to raise any questions they wanted to and to criticize the items Objectionable words or phrases were pointed out, a few unrealistic answers were criticized, and better substitutes were found

Further revision of the test results from the above procedures, and the test is reproduced for the *collection of data* on a larger scale The actual number varies with the facilities for trial testing, but is normally large enough to make possible the establishment of time limits, the checking of the clarity and completeness of directions, the locating of ambiguous or offensive items, and the analysis of the internal consistency of the test The subjects at this stage should be a sample of those for whom the test is designed, not only because different types of groups may require different amounts of time or need directions which go into varying amounts of detail, but also because items that work well with one type of subject may not work well with another for example, a question may be well-phrased and have a right answer for unsophisticated subjects, but may be unanswerable by more sophisticated examinees because of oversimplification of matters which they know to be complex

An *analysis of the internal consistency* of some tests is not possible at this stage, either because some apparatus tests with time scores have no items or parts, or because the test may not be scorable until it has been item-validated If, as is generally the case with aptitude tests, there is

an *a priori* method of scoring based on right and wrong answers, this scoring key needs to be analyzed to make sure that answers keyed as "right" are in fact generally chosen by those who make high total scores, and that the wrong answers are more frequently chosen by those whose total scores are low The test is then *revised again*, in order to eliminate poor items and sharpen those that are ambiguous, after which it should be ready for large scale administration

Standardization

The principal problem in administering vocational tests for standardization and validation is *whom* to test, and at what stage of their careers. The question of *how many* is more easily answered at least in theory Whether the test is to be used in guidance or in selection (in which this writer includes placement and promotion unless otherwise specified), it is obvious that it should be standardized on persons for whom the chosen criterion or criteria of success are or will be available But this raises a problem which has plagued psychologists since the beginnings of aptitude testing, for if the test is standardized on a group who are already employed in the occupation, and for whom criterion data are presumably readily procurable, there will be a real question as to the value of the test when used with persons who have not yet entered the field Specifically, will a low score made by a high school or college student indicate a relative lack of the aptitude measured, or will it reflect primarily what is already known, namely, his lack of training and experience in the field in question? If, on the other hand, the test is administered to students or others who have not yet entered the field in question, how is one to validate it? The lag between testing time and that at which criteria of success become available may be considerable, and the loss of cases through entry into other fields not being investigated and through change of address is certain to be almost prohibitive

Longitudinal validation studies of the type just mentioned are rare Strong's studies of his Vocational Interest Blank have generally employed the *ex post facto* validation of differentiation between people employed in various occupations (775 Ch 7), but he has also administered his inventory to miscellaneous college students and followed them up about ten years later (775 Ch 16) in order to ascertain the relationships between their test scores on the one hand and entry into and stability in various occupations on the other Longitudinal validation has been used more in selection programs, especially those involving training after pre-

liminary selection The Armed Forces frequently selected on the basis of tests which were first validated by giving them as though for use in selection and then checking their results against success in training, schools of nursing, medicine, engineering, and other professions do likewise, although in these cases there is no guarantee of employment if training is completed As users of tests in personnel work become more test sophisticated, as users of tests in guidance become more exacting in their requirements, and as constructors of tests raise their standards through familiarity with good practices, longitudinal validity studies should become more numerous

In the meantime cross-sectional validation studies are the commonly available type Strong first validated his inventory by contrasting the answers of men in one field with those of men in other fields, Kuder is now doing the same with his, although the first validation was by internal consistency (802), the numerous sets of norms compiled by the Minnesota Employment Stabilization Research Institute compare workers in one field with those in others or with the general population (589), the material comprising the bulk of this book deals with *group differences* and relationship to *success in training*, rather than with *success in an occupation*, because of this emphasis in the research It may be well to point out, however, that the result may not be as disastrous for vocational counseling as one might suppose, for work by Strong and by Carter (145), the most complete along these lines, shows that the results of some *ex post facto* validated tests can legitimately be applied to untrained and inexperienced persons if one knows what corrections to make for maturation This finding for Strong's Blank has been confirmed in other ways with other tests, for example, by determining the effects of training and of age on the Minnesota Clerical Test (see Ch 8)

The number of cases to be obtained, it has been stated, is more readily decided upon than whom to test and when to test them, but is inextricably involved in both of these The determining factors are the number needed in order to compute certain statistics and the number that can, in a given practical situation, be tested If the test is being standardized for selection in a department which employs 200 workers in one job and hires fifty new people each year, and if results are to be available for use in a reasonable period of time, it is clear that pre-selection testing and validation cannot be based on more than 50 or 100 cases, and that validation upon persons already working is not likely to be feasible with more than 250 or 300 cases As these numbers are large enough for computing

correlation coefficients and critical ratios, test construction and validation may well be worth while in this situation Certainly the sample would be adequate if the test is to be used only to select for that job in that concern providing labor market and job remain the same, as it includes the whole universe in question rather than just a sample

If the test is to be standardized for counseling in connection with the choice of an occupation the problems of numbers and sampling become much more acute While it is relatively easy to make sure that a job in one factory is in fact *one* job rather than a number of *different* jobs, making sure that the persons who are nominally engaged in a given occupation are in reality doing the same type of work is almost impossible, for if they are to be a good sample they must be distributed throughout the country and analysis of their work is likely to be impossible The test constructor has then to content himself with other devices which may help him select a well-defined and homogeneous group He may, like Paterson and his associates (588) confine his study to a thoroughly studied and well-defined group of boys in one junior high school in one community, he may follow their lead in a series of other studies (589), and select a cross section of the employed population of one city which is distributed among the major occupations in the same manner as the employed population of the United States as a whole Both groups may then number only in the hundreds, being well selected But in the former case, the counselor must assume that success in mechanical activities will be judged in the same way in his school or community as in Paterson's, and that the same psychological and social factors operate in his subjects in approximately the same way, or he must refuse to use the test without a local validation study of his own In the latter case he must assume that stenographers and typists in Minneapolis do the same types of work, requiring the same *types* and *degrees* of aptitudes and skills, as the stenographers in his own community, and similarly with retail salesmen, garage mechanics, policemen, etc, or he must refrain from using the tests until he has gathered his own norms and his own validation data The assumption may be quite sound in some instances and quite unsound in others, the writer suspects that it may be true for bank tellers, but false for retail sales-clerks Observational evidence for the latter assumption lies in the differences between standards for clerks in dime stores and in more expensive establishments, which govern both the referral of girls to such stores by placement workers and their selection by employment managers

The solutions to the sampling problem used by Strong, who faced it repeatedly and was primarily interested in the counseling values of his test, followed no uniform pattern and illustrates the opportunism which problems of time, money, and co-operation have forced upon test construction workers The psychologists on whom Strong standardized his psychologist key constituted more than one third of the full members of the American Psychological Association at the time of standardization (some 200), and were scattered throughout the country, having been reached through the membership list of the Association This would seem to be a good sample of academic psychologists, although it may have slighted applied psychologists, some of whom were not members of the Association On the other hand, the group upon whom the key for social science teacher was standardized consisted of more than 200 teachers employed in the state of Minnesota They may have been a good sample of such teachers in that state, but there is no way of knowing whether they were also typical of social science teachers in New Hampshire with its rather different population, in Georgia with its different culture and salary standards, or in other states and localities Obviously, the counselor using such a test needs to know the characteristics of the population on which it was validated, and the extent to which the latter resembles the population with which he is working, before he can draw any legitimate conclusions from its scores It is, therefore, important for the test constructor to choose his validation group well and to describe it in detail

Validation

The terms standardization and validation have been used synonymously in the preceding section, because the standardization of a vocational test implies collecting data which make possible validation If the test is administered to persons with whom its use is appropriate, norms are gathered, and its significance ascertained, much of the process of validation is already accomplished in standardization In the sense in which the term is used in the sequence of steps outlined here, validation is therefore the statistical procedure of analyzing test results in relation to criterion data (see page 651 for definition) In work with some types of tests this process consists of just one step, the determination of the relationship of test scores to the criterion, in work with other types of tests, however, it involves another step before scores can be validated, specifically, the validation of each item in the test

Item validation, the determination of the extent to which a given

question is answered one way by the "success" and other ways by the "failure" group, impresses the novice as a laborious procedure It is this, but it frequently proves its worth and is often indispensable to test construction For example, the writer and three associates developed a personality inventory, referred to previously, for use in aviation cadet selection and classification The items had no inherently right or wrong answers, as they dealt with satisfaction and dissatisfactions in such things as drill, strafing ground troops, bombing towns and cities, and being an officer, but the item writers naturally had hypotheses concerning the psychological soundness of the attitudes expressed, and of the possible significance of these reactions for success in flying training One of the collaborators (John L Wallen) constructed two *a priori* keys for this inventory, one of them intended as a measure of morale, the other of atypicality of attitudes and behavior The former was strictly *a priori*, but the other contributors to the inventory (Robert R Blake and Joseph Weitz) agreed that the responses scored as indices of poor morale would, in fact, be considered symptoms of poor morale by most competent judges The atypicality key was more objective, in that it was empirically derived all responses chosen by small percentages of the cadets in the standardization group before the predictive value of the test was known were weighted in the atypicality key One of the collaborators (Blake), while agreeing with the logic of each step in the construction of these keys, was convinced that they would not have any validity for success in aircrew training, the others, though pragmatic in their attitudes, thought they might prove valid When criterion data in the form of graduation-elimination reports arrived from primary flying schools and scores on the two *a priori keys* were validated against them, the scoring keys were found to have validities of approximately zero (801) The next step was therefore to validate each item against pass-fail in flying training, when this was done, quite a number of items were found to be answered predominantly in one way by the successes and in other ways by the failures A new, empirical key was therefore made and cross-validated on another part of the sample not used in the item validation it proved to have a validity of about 20 (significant at the 1 percent level) While this was not very high, the test was unique enough for its contribution to the cadet classification battery to raise the latter's validity from about 66 to about 69, an improvement easily worth twenty minutes of testing time and a moment of scoring (316 736–746)

This example brings out clearly the importance of item validation in

tests and inventories which have no inherently right or wrong answers, for even the best logic often fails in constructing vocational tests Even when a test has right and wrong answers, however, the right answer is not necessarily the best for persons in a given occupation If, for example, being well informed on the hobby of philately were characteristic of men who succeed in pilot training, being able to select the correct definition of the term "wove" from among four false definitions would be a "right" answer for potential pilots, but, if knowing about stamps and stamp collecting were characteristic of men who *fail* in pilot training, the correct definition of the term would be a "wrong" answer for pilots If the latter were the case (the example is fictitious) a test of philatelic knowledge might be validated as a test, without item validation, but one would need to be certain that it was philatelic knowledge as such that was prognostic of failure, and not just knowledge of certain aspects of stamp collecting such as the technicalities of paper-making, colors, and perforations, as contrasted with the historical and geographic knowledge which a careful stamp collector also acquires Hence the usefulness of item analysis, as described by Davis (196) This problem does not arise when the test is of a clearly homogeneous type, for example, a spatial visualization test utilizing two-dimensional forms in each item, for in their case both logical analysis and internal consistency indices demonstrate the fact that what is measured by one item is also measured by other items

The *validation of scores* is generally done by correlating the score made on the test with the criterion data Thus the validation of a test of ability to judge spatial relations for military pilots involved computing biserial correlation coefficients for test scores and pass-fail reports of cadets who entered primary school after taking the test, and the validation of Strong's life insurance salesman's key for success in selling life insurance involved the correlation of dollar volume of sales with test scores, using the product moment method (772) In many cases other methods are used, the principal reliance in Strong's insurance study, for example, being placed on the analysis of the percentages of men with a given letter grade on the interest inventory selling a given amount of insurance (enough to make a living as a salesman) This method is known as the percent of overlapping technique using a cut-off score, and differs only superficially from a third, in which group differences are expressed by means of a critical ratio These are standard techniques described in detail in elementary texts on statistics

The choice of method is dictated by the form in which the data are expressed reports concerning having passed or failed a course cannot be used in computing Pearsonian correlation coefficients, but do lend themselves to the use of biserial r's Data like Strong's lent themselves to either correlation or percent-of-overlap analysis, although he used both procedures, more emphasis is placed on the latter technique, because the nature of interest scores makes letter grades more meaningful than standard scores (775 67) and because the fact of earning or not earning enough money to live on seems more important in judging success as a salesman than differences above or below that amount

Cross-Validation

It has long been an accepted principle of test construction that a test should be not only validated, but cross-validated, that is, administered to another comparable group and scored in the same way, to ascertain whether the validity for the second group is as high as for the first This need was brought out by the fact that validities in subsequent studies were often lower than those in the original study of a test, as a result of special factors present in the criterion group which are not present in the cross-validation groups These factors operate especially in small samples in which, for example, a disproportionate number of members may, as a result of pure chance or of administrative bias, come from one part of the country, be younger than the occupational universe from which they are drawn, or have some other things in common which are not so common in other samples of the same occupational group

A good illustration of the operation of this type of regression toward the mean is found in the author's study of avocational interests (791 60), in which scoring keys for the hobbies of model-train building, instrumental music, photography, and stamp collecting were found to regress from mean standard scores of 50 for the criterion groups to means of 36, 24, 33, and 26 for the respective cross-validation groups When expressed in terms of group differentiation, these results meant that, although the scoring keys differentiated quite well between the criterion groups on which they were based, they failed to differentiate other similar hobby groups in the case of the philatelic key, failed for all practical purposes in the case of photography, and differentiated somewhat in the case of model engineers and fairly well only in the case of amateur musicians Strong (775 637 ff) studied this problem, and found that, although

groups could sometimes be differentiated with as few as 50 or 100 cases in the criterion group, better differentiation was obtained, with minimal regression toward the mean in cross-validation, when criterion groups of from 250 to 500 are involved

Although the need for cross-validation has been recognized in the literature it has in fact too often been honored in its breach because of practical reasons such as time, money, and the difficulty of obtaining co-operation from sufficiently large groups Some dramatic instances of reversed relationships in cross-validation are reproduced from Stead and Shartle (750) in Figures 4 and 5 (pp 169 and 170)

The experiences of psychologists in World War II have again driven home the fact that cross-validation is essential, despite Strong's conclusion (775 650) that, when a large criterion or original validation group is used (and additional cases are difficult to obtain), cross-validation may be dispensed with Experience repeatedly showed that a test validated on several hundred aviation cadets might appear valid until evidence was obtained on another sample, at which time it would lose all semblance of validity In one study the Rorschach Psychodiagnostic was administered to cadets, and ratings of their probability of success in training were made by trained examiners who were also somewhat familiar with the requirements of flying training (316 625–637) In the validation or criterion group consisting of every other tested cadet (N=283) the biserial r with pass-fail was 23, the standard error being 09 When the cross-validation was completed on the other half of the tested group the correlation fell to approximately zero The original figure was not very high, it is true, but with a battery of tests which occupied one and one-half days of the cadet's time and had a validity of 66, each research test which had a validity of 20 and a low correlation with the tests actually in use was carefully scrutinized as a potential contributor to the battery, and a number of such were found which repeatedly yielded validity coefficients of about the same size and added 03 or 04 to the validity of the battery

The techniques of cross-validation are the same as those of validation with the original group Sometimes they are applied after a second round of testing and the collection of new cases, but more commonly it is found more practical to gather enough data at first to carry out both procedures, doing the validation on even-numbered cases, for example, and the cross-validation on odd-numbered cases This insures controlling the effect of the times at which data are obtained, and yet provides two groups for study

Factor Analysis and Factor Validation

A further step in test construction and validation has been added by Guilford (316 Ch 28, 317,319), through the application of factor analysis to test construction in personnel selection Briefly, it consists of analyzing tests in order to ascertain their factorial composition, and of analyzing the criterion in order to determine the nature and weight of the factors which enter into it The former step makes possible the refinement of tests, to make them factorially pure, this has the advantage of cutting down the number of tests needed to predict success, by eliminating overlapping of tests and making each test do a maximum of work The latter step, analysis of the criterion, indicates what types of tests should be stressed in order to improve predictions Illustrations of each of these procedures follow, again taken from aviation psychology because the most extensive applications to date were made in the Army Air Forces

Factorial Analysis of Tests The use of factor analysis implies that tests can be statistically analyzed into a limited number of underlying traits or aptitudes, or, conversely, that existing tests actually measure a number of traits which can be isolated by statistical analysis To attempt to describe the procedures of factor analysis would be out-of-place in this text, but some understanding of the significance of factor analysis for test construction and validation is in order The application of the Thurstone centroid method of factor analysis with rotation of axes (839) to a battery of tests results in the isolation of three types of variances or components 1) several common factors, that is, components which appear in several tests, 2) possible specific factors, appearing in only one test, and 3) error variance, arising from the unreliability of the measures These common factors, having been arrived at by a process which is largely mathematical, may or may not make psychological sense, it is by rotating the axes that meaningful factors are made to emerge This is a somewhat subjective procedure, calling for judgments on the part of the statistician Even more subjective is the naming of the factors that have been isolated, this is done by inspection of the kinds of tests which are saturated or loaded with a given factor, to ascertain what the common elements seem to involve

Guilford (317) provides an illustration of how factor analysis can help one better to understand what tests are measuring, graphically presented in Figure 2

This figure shows the proportions of factor variances in three of the

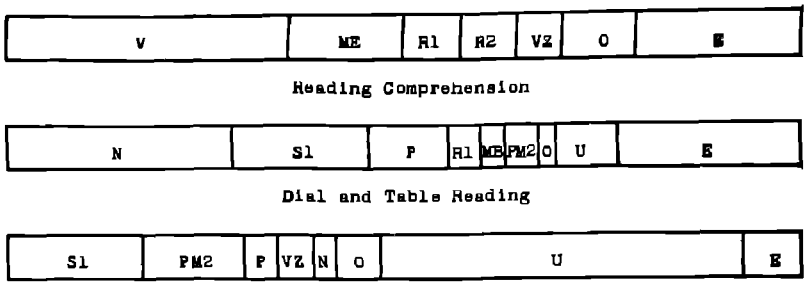


FIGURE 2

FACTORS MEASURED BY TWO PAPER AND PENCIL TESTS AND ONE APPARATUS TEST, ADMINISTERED TO AVIATION CADETS

Illustrating the complexity of simple tests and the unknown quantities in miniature situation tests. After Guilford (317)

tests used in the AAF Aviation Psychology Program. These tests were developed in the standard ways already described. That is, it was thought that reading comprehension might play some part in flying success, so a Reading Comprehension Test was developed with aviation types of materials. Pilots, navigators, and bombardiers make much use of books of tables and take many readings from dials. A Dial and Table Reading Test was therefore developed, using dials such as those in airplanes and tables such as are used in navigation. Reaction time is frequently mentioned by pilots as an important characteristic in flying, quick response to a variety of stimuli being obviously important in taking off, landing, and in many emergencies, hence a Discrimination Reaction Time Test was constructed, along lines long used in laboratory studies in physiological psychology.

The job analysis procedures used in developing these tests were obviously those of observation and deduction. The tests were, in the cases of Reading Comprehension and Discrimination Reaction Time, attempts to measure more or less unitary traits, and, in the case of Dial and Table Reading, an attempt to duplicate the job situation in miniature.

As Figure 2 brings out, all three tests were complex in their factorial composition. This was true not only of the miniature situation test,

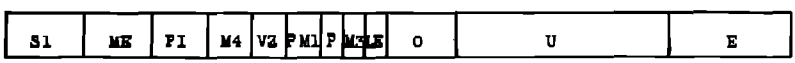
which might have been expected to draw on a variety of abilities, but also of the two tests which are normally thought of as being simple in their composition The reading test draws on the following abilities verbal comprehension, mechanical experience (some of the content was mechanical), general reasoning, analogic reasoning, visualization, and several much less important factors The discrimination reaction time test requires ability to judge spatial relations, psychomotor precision, perceptual speed, visualization, numerical ability, several minor factors, and a relatively large number of unknown factors, one of which might be reaction time The dial and table test measures six major factors (number, spatial relations, perceptual speed, general reasoning, mathematical experience, and psychomotor precision), a few minor factors, and some unknown factors Such unknown factors, if not specific to the test, emerge because the test battery does not include enough other tests for them to be clearly recognizable

These three tests were found to measure, not three traits, but a total of eleven Of these, six are measured by more than one test This is clearly not economical, as one good measure of a given factor would be less time-consuming than three tests It is also inefficient from the point of view of prediction, as the validity of one test may be due to one of the factors it measures, whereas others that it also taps may actually tend to lower its validity, as when they correlate negatively with the criterion In such a case positively and negatively significant factors tend to counteract and cancel each other in the same test

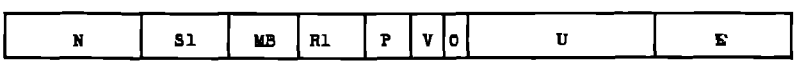
The contribution of factor analysis to test construction is, therefore, to make possible the refinement and purification of tests, and to reveal what kinds of tests may actually be developed The three tests just described yielded ideas for eleven different tests, some of which might be positively significant for pilot selection, some negatively, and some not at all The construction of eleven separate tests makes possible the differential measurement of these eleven traits, and improves predictions based on the validities of these traits Profiles showing the scores on independent traits such as these are much more useful in counseling a person, provided the validity of the traits measured is known Guilford's unique contribution lies in his having not only isolated the underlying factors of this extensive battery of tests, but in having ascertained the significance of these factors for success in several occupations This latter topic is expanded in the next paragraphs

Factorial Analysis of the Criterion When factor analysis is applied to the criterion of success, two major types of results are accomplished.

First, the occupational significance of the factors is made clear, permitting the counseling of individuals on the basis of factor profiles or the weighting of factors rather than of tests in selection programs As is pointed out in the discussion of the Primary Mental Abilities Tests (Ch 6), the drawback of factorially pure tests has been the lack of evidence to guide the interpretation of their results The second outcome is a better understanding of what it is one is trying to predict, that is, of the nature of success in the occupation in question Factor analysis of the criterion gives one an objective description of what it is that is being predicted, to supplement the observational data of traditional job analysis and the deductions from test validities But the very nature of factor analysis imposes some limitations of a very serious nature on the second type of use of the technique with the criterion As many writers have pointed out, one can extract from a factor analysis only that which is put into it More concretely, the only factors which can be isolated are those which are tapped by more than one test in a battery If, therefore, the battery of tests used in the analysis is limited in scope and fails to include some traits which might be measured (and all batteries are more or less open to this criticism), the analysis of the criterion will leave undescribed some of the abilities which it requires An indication of the extent of these unmeasured components is of course, provided by the unknown-factor variances



Pilot Criterion



Navigator Criterion

V—verbal comprehension
ME—mechanical experience
R1—reasoning I (general)
V2—visualization
N—numerical
S1—space I (spatial relations)
P—perceptual speed
MB—mathematical background
FI—pilot interest
M4—memory IV (content)
M3—memory III (picture symbol)
PM1—psychomotor coordination
LE—length estimation
O—other minor factors
U—unknown or specific
E—error variance

FIGURE 3

FACTORS IN PILOT AND NAVIGATOR CRITERIA

As revealed by Army Air Forces factor analyses of success in training
After Guilford (317)

Figure 3, also taken from Guilford (317), shows two criteria analyzed in the same way as the three tests already discussed The pilot criterion was found to be composed of 27 common factors, about 52 percent of the variance of success or failure in pilot training could be accounted for by 23 of these factors If other tests of appropriate but unknown types had been included in the battery, another 28 percent of the variance could perhaps have been predicted, leaving 20 percent of the variance in success or failure due to lack of reliability Nine known factors accounted for about 56 percent of the navigator criterion, apparently success in navigation training was more easily predicted, and less complex in nature, than was success in flying training

It is interesting to note that success in pilot and in navigator training have little in common, according to the data in Figure 3 only spatial relations and perceptual speed appear in both occupations This is in contrast with the three tests for which factorial data were presented, and which overlapped more completely in their components despite superficial differences and some unique factors

What these data make clear for vocational counseling is that number ability is not important in success in pilot training, and need not receive attention in profile interpretation, that mechanical experience, visualization, and psychomotor precision (among other abilities) differentiate pilots from navigators, and that navigators, on the other hand, are helped by the possession of number ability and mathematical background These facts are brought out more clearly by factor analysis of the criterion than they could be, for example, by an analysis of the differential validities of impure tests such as that for reaction time or reading comprehension

For improvement of predictions of success in pilot and navigator training these data make clear the facts that there is still considerable room for improvement in the test battery and for the development of tests measuring other factors We have already seen that approximately 28 percent of the variance in pilot success could still be predicted if suitable tests were available The graph shows that there is probably less room for improvement in navigator selection But just how this improvement is to be effected, or what types of traits should be tested, is not made clear In order to get clues as to what these traits are, one must still depend on the traditional type of job analysis, whether for the selection of existing tests or for the devising of new instruments for inclusion in the battery and in the next factor analysis

CHAPTER IV

THE NATURE OF APTITUDES AND APTITUDE TESTS

Definitions

THE term "aptitude" is generally used loosely both by laymen and by vocational psychologists and counselors Its meaning varies not merely from one user to another, but even from one time to the next in the speaking or writing of a given psychologist or educator It is used in either of two ways, as when we say that a man has a great deal of aptitude for art, meaning that he has in a high degree many of the characteristics which make for success in artistic activities, or when we say that a person lacks spatial aptitude, meaning that he lacks this one specialized aptitude which is of varying importance in a number of different occupations In the former instance the word is used not to denote a unitary trait, nor even an entity of any sort, but rather a combination of traits and abilities which result in a person's being qualified for some type of occupation or activity In the latter case the word "aptitude" is intended to convey the idea of a discrete, unitary characteristic which is important, in varying degrees, in a variety of occupations and activities

These two different meanings have been attached to the term as a result of the tendency of psychology to use existing words which already have popular meanings, redefining them in the process for the sake of clear thinking, instead of coining new terms of Latin or Greek origin as is done in fields such as biology and physics Both the popular concept of *aptitude for a vocation* and the scientific concept of *aptitude important in vocations* are essential, it is important, however, that the meaning intended be clear In general, counselors and personnel men tend to think in terms of vocations and jobs, and therefore to use the term in the broad popular sense, while psychologists tend to think in terms of individual differences and traits, and therefore to use the term in the narrow scientific sense As most of the literature on tests is written by psychologists,

and most of the tests were constructed by psychologists, the counselor or personnel man needs to develop the habit of starting with the narrow scientific meaning of the term and of translating the psychological trait or characteristic into broader vocational terms Similarly the psychologist, if his report of test results is to be meaningful and useful to the counselor, social worker, personnel man, or teacher, must be able to translate trait data into vocations

Various combinations of traits and abilities may make for success in a given field One teacher, for example, may be successful because of scholarly ability, interest in his subject, and a desire to share it with others which result in a clarity of presentation, a wealth of material, and a warmth of manner which more than make up for a relative lack of interest in people as individuals and a dislike of the routines and details of classroom management Another teacher may be equally successful because of his genuine interest in students, his warm and friendly manner, and his skill in classroom management, even though his scholarship and academic ability are mediocre Similiar differences could be pointed out among successful lawyers, salesmen, foremen, assembly-line workers, and probably even machinists and draftsmen, although the facts are not so clear in the case of the skilled trades and lower technical occupations

Because of the varying combinations of special aptitudes and traits which make for success in a given occupation, it is desirable to continue the scientific use of the word aptitude in testing and test research For this reason, the term will be used in its narrower sense in this book, except when expressly defined otherwise, as in the phrase "aptitude for the medical profession"

Even in its narrower scientific sense, however, the word aptitude is by no means consistently and clearly used in the literature on tests In Warren's *Dictionary of Psychology* (910) it is defined as a *condition* or *set of characteristics* indicative of ability to learn This implies that an aptitude is not necessarily an entity, but rather a constellation of entities, the set of characteristics which enables one person to learn something may even be different from that which enables another person to learn the same thing, in this case, we arrive back at the popular definition Bingham (94 16–18) uses approximately the same definition, further confusing the picture by adding a readiness to develop interest in using the ability In some unpublished material Seashore and Van Dusen have attempted to define the term more rigidly, saying that an aptitude is a measure of the probable rate of learning, which results in interest and

satisfaction, and is relatively specific and narrow The scientific study of an aptitude or of any other entity requires that one be able to name it (whether meaningfully or by means of a symbol such as x), describe it, and locate it in a variety of individuals and situations This means that it must be relatively constant in its nature and composition Warren's and Bingham's definitions are therefore useless to a scientist or to a counselor, while that proposed by Seashore and Van Dusen is more useful in that it prescribes narrowness and specificity Accordingly, a scientific definition of aptitude would provide for *specificity, unitary composition,* and the *facilitation of learning* of some activity or type of activity

In practice, the requirement of unitary nature is frequently minimized The Minnesota Vocational Test for Clerical Workers or number- and name-checking test is, for example, a test of about as simple an entity as one could expect to find, and yet factor analysis shows that the names test includes not just a speed and accuracy of discrimination factor identical with that in the numbers test, but also an intelligence factor not found to any appreciable degree in the numbers test (21) The Bennett Mechanical Comprehension Test, and others like it, are generally assumed to measure a special aptitude and yet the best available evidence suggests that mechanical information and ability to visualize space relations play major parts in it (see below, p 221) In our present state of knowledge and with the current refinement of our techniques, it seems wiser to be satisfied if the aptitudes measured are relatively distinct and have some validity, than to devote too much time to obtaining pure traits The quick success of this global approach in Binet's work with intelligence tests, discussed in Pintner (604 Ch 2), has been borne out in aptitude studies such as the Minnesota Mechanical Abilities Project (588) and in interest research such as Strong's (775), and even more recently in the slow rate of progress which has characterized the pure trait approach as used in Thurstone's work on primary mental abilities (838) and Kuder's work on primary interests (446) In Thurstone's work the development of sufficiently refined and reliable instruments has been time-consuming and the results in terms of educational or vocational validity disappointing (see below, p 141), and in Kuder's it has taken thirteen years to develop an instrument with vocational significance (see below, pp 445, and 459) This is not to decry the importance of such studies of primary abilities and interests, nor of the resulting tests, on the contrary, they undoubtedly are the beginning of a new era in aptitude and interest measurement and foreshadow tests which are more refined

and more valid than any we now have Guilford's work (316 Ch. 28, 317, 319) has demonstrated this But for most practical purposes it is still true that the best *available* tests are those which do not over-stress the unitary nature and purity of the aptitudes or traits measured

A fourth and final characteristic of an aptitude should probably be added to our definition, namely, that it is relatively *constant* If behavior or success is to be predicted, the entity upon which the prediction is based should be relatively stable An aptitude which varied irrationally from one day, month, or year to the next would not provide a sound basis for predicting achievement at some future date To put it statistically, an aptitude which is itself unreliable could be neither reliably measured nor significantly correlated with anything else. This question of the constancy of traits has, as the literature of recent years makes amply evident (294,501,760,832,917,918), been a prime source of disagreement among psychologists The attending controversies are too involved for adequate discussion to be possible here It seems wiser to side-step detailed discussion and simply to state the author's conclusion that, whether largely innate or largely acquired, the aptitudes about which we know something appear to become crystalized in early childhood and that after that they are relatively constant They may then perhaps be affected by especially drastic or traumatic experiences, but can otherwise be thought of as not being appreciably affected by education, special training, or experience This is not to imply, however, that specific practice on the items or materials of the test itself will not, through practice effect, raise the subject's test score, the contrary is true, but that does not indicate a change in the degree of aptitude As demonstrated in a number of different studies, interests and personality traits are crystalized later than aptitudes, in adolescence (144,568,771,775) The evidence for specific aptitudes and traits will be viewed later, as each test and the work done with it is studied in detail

Two other terms need brief definition One of these is the word *skill*. It is used here, and in most discussions of abilities, as synonymous with *proficiency*, to denote the degree of mastery already acquired in an activity Thus a typing test is a test of skill, and a trade test is a test of proficiency The other term is *ability*, which Bingham (94 19) uses to denote either aptitude or proficiency or both, leaving it to the context to indicate the meaning, and which Seashore and Van Dusen prefer to use as a synonym for proficiency but not for aptitude In view of the convenience of having a general term the writer prefers to use ability to include both

aptitude and proficiency, using one of the latter terms when clarity and specificity require The term *trait,* it might be noted, is used as comparable, in the field of interest and personality, with the term *aptitude* in the field of abilities

The Basic Aptitudes

E L Thorndike once suggested that there are probably three types of intelligence abstract, mechanical, and social Since that time there has been a great deal of speculation and research on the nature and number of special aptitudes T L Kelley used factor analysis and a variety of tests in order to study the question (418), concluding from his data that aptitudes may be classified as verbal, numerical, spatial, motor, musical, social, and mechanical He provided also, in his scheme, for various types of interests Spearman made another analysis (731), using other tests and a quite different method of factor analysis, since then he and his students in England have modified and elaborated his position, concluding that there are one general or intelligence factor "*g*," a number of group factors such as word fluency, perseveration, and goodness of character, and many specific factors which are found only in one test or situation Thurstone's work (838,839) in factor analysis and the organization of special aptitudes has probably had more influence in America than has any other Using the centroid method of factor analysis he isolated the following special aptitudes number, visualization, memory, word fluency, verbal relations, perceptual speed, and induction This research has borne fruit in the Chicago Tests of Primary Mental Abilities (see pp 132 ff), which measure six factors, number, verbal meaning, space, word fluency, reasoning and memory

Two other factor analyses of aptitudes have followed Thurstone's, each of them using a greater variety of tests and therefore isolating more factors than its predecessor One of these was made by the United States Employment Service, under the direction of Shartle (735), and the other by the Army Air Forces, under Guilford's supervision (316,317) The lists of factors arrived at by each of these investigators are combined in Table 2, in order to show how the list of presumably unitary human abilities lengthens as the investigations become more thorough-going It will be noted, for example, that what Thurstone thought was one single aptitude, perceptual speed, was broken down into two factors, perception of symbols and perception of spatial forms, in the USES study, and into two apparently similar aptitudes in the AAF investigation What Thurs-

tone's study isolated as one factor, memory span, did not appear at all in the USES research because no memory tests were used, but was broken down into three distinct types of memory factors in the AAF analysis As might be expected in the case of a program which devoted

TABLE 2

THE EXPANDING LIST OF PRIMARY ABILITIES
According to Thurstone (839), Shartle (735), and Guilford (316)

Thurstone 1938	*USES (Shartle 1945)*	*A A F (Guilford 1947)*
Spatial	Spatial	Spatial Relations I
		Spatial Relations II (Right-Left Discrimination)
		Spatial Relations III (Unknown)
		Visualization
		Mechanical Experience
Perceptual Speed	Symbol Perception	Perceptual Speed
	Spatial Perception	Length Estimation
Number	Numerical	Numerical
		Mathematical Background
Verbal Relations	Verbal	Verbal
Word Forms		
Memory Span		Paired Associates Memory
		Visual Memory
		Picture-Word Memory
Induction		
	Intelligence	General Reasoning
Reasoning or Deduction	Logic	Analogic Reasoning
		Sequential Reasoning
		Judgment
		Planning
		Simple Integration
		Complex Integration
		Adaptive Integration
	Speed	
		Psychomotor Speed
	Aiming	
		Psychomotor co-ordination
	Finger Dexterity	Psychomotor Precision
	Manual Dexterity	
		Kinesthesis
		Carefulness
		Pilot Interest (Active-Masculine)
		Social Science Background

considerable time and talent to the development of new types of tests, the Aviation Psychology Program battery revealed, when analyzed, far more primary traits than were isolated by the other investigations Thurstone's list included only eight factors, Shartle's 11, and Guilford's as many as 28 The list will no doubt continue to grow, as evidenced by the 28 per-

cent of the variance in pilot success which was not, but might be, predicted, if suitable tests were available

Other factors which may in time be isolated and added to our list of human abilities are suggested by Seashore's (690) and Meier's (519) studies of musical and artistic ability, discussed in a subsequent chapter In the meantime, the lists in Table 2 provide a good basis for job analysis and test selection or construction

Thurstone's method of factor analysis provides for the isolation of independent factors or aptitudes For this reason, most of the aptitudes named above are relatively independent of each other Some, such as those normally included in the concept of general intelligence, are more closely related, but the intercorrelations are still lower than reliability coefficients, that is, too low to make a test of one aptitude or factor a good index of the score on the test of another factor Tests of spatial visualization frequently have moderately high correlations with tests of intelligence, but this is an artifact arising from one or both of two causes, depending on the circumstances first, tests of intelligence often include tests of spatial judgment (e g , Army Alpha and the Army General Classification Test), and secondly, as Garrett has recently pointed out (281), this and other factors which appear to constitute intelligence in children become differentiated with increasing maturity and constitute, in reality, special aptitudes rather than aspects of general ability Because their rates of maturation are similar, the more abstract abilities appear to be more closely related to each other than the more concrete abilities Tests of manual dexterities, not included in most factor analysis studies, have been analyzed to show that the concrete abilities which they measure are more discrete and have lower intercorrelations than do the more abstract aptitudes, this will be seen in the chapter on tests of manual dexterities

Despite the demonstrated independence of special aptitudes, there is a tendency for groups of people who score high on a measure of "general aptitude" to make good scores on other tests, whether of special aptitudes or of personality traits As Terman pointed out in his "Genetic Studies of Genius" (819), the good things tend to go together, a statement amply borne out by varied psychological and social data on more than one thousand gifted children who were followed into adulthood (823) It is therefore not surprising that, in counseling practice, one encounters persons who make high scores on tests of academic aptitude and on almost any other test one administers to them, and others who not only make low scores on tests of general mental ability, but distress one by also making

low scores on any other instrument which is used in the search for some "hidden talent" which might be capitalized and built upon It is well not to be overimpressed by such cases, however, as it has been demonstrated (603) that they are outnumbered by those whose aptitudes and personality traits vary considerably, giving them some assets and some liabilities

Methods of Measurement

The most valid method of measuring an aptitude, that is, a unitary factor in the ability to learn something, would be to find out what part of the activity or skill to be learned is most heavily saturated with that factor, have the subject learn it, and compare his rate of learning with that of other persons with comparable backgrounds This is in most cases an inordinately expensive method, although selection on the basis of success or failure in an initial learning period is still the method used by many colleges and professional schools which consciously admit two or three times as many beginning students as they expect to graduate and flunk those who make the lowest grades during the first year It is the method that was used in selecting cadets for pilot training both in the AAF and in the RAF prior to the development of adequate psychological tests, and it is that used by many businesses and industries even now despite the interest of many in taking advantage of the possibilities of scientific personnel selection and the great strides made in this direction by some life insurance companies, manufacturing concerns, banks, and retail establishments Experience as well as theory has demonstrated that it is less expensive and better policy in other ways to analyze the task in which success is to be predicted, develop and validate tests for predicting achievement in that task, and select on the basis of test and other personal data than to do a less careful job of initial screening and depend more on selection on the job In the same way, it is less expensive, less discouraging, and less difficult for a high school or college student, unemployed man or woman, or adult considering a transfer or change of work, to take a series of tests and analyze his experiences in order better to ascertain his ability to learn a new task or to adjust to an occupation than it is for him to try it out as a probationist or actual employee

There are different types of tests of aptitudes, each of which has its disadvantages as well as its advantages The user of tests, as well as the test constructor, should be familiar with these They will be briefly described here in terms of contrasting types or dichotomies

Miniature tests may be contrasted with tests of *abstract traits* or aptitudes In the former, the task in which learning or success is to be predicted is reproduced in miniature and perhaps simplified form, as, for example, in the familiar lathe-type or two-hand spatial judgment and co-ordination test This miniature test, used successfully in selecting shop students, duplicates on a smaller scale both the apparatus and the arm and hand movements of a lathe In the test of abstract aptitudes the job has been analyzed and one or more of its essential characteristics has been abstracted and put into test form Thus in the MacQuarrie Test of Mechanical Ability there are a series of tests of eye-hand co-ordination and of spatial judgment, one of which involves tapping three times in each of a series of small circles, another tracing a line through the variously placed small apertures in a series of barrier lines, and still another judging the number of blocks touching others in a series of piles of blocks In this case the test bears no superficial resemblance to the original task or activity, let us say lathe operation, but some of the essential aptitudes seem to be measured

The miniature type of test has a number of advantages Its face validity or obvious similarity to the task in question makes it appeal to the examinee who is interested in such work Being a small scale task, it is very likely to involve the same aptitudes and skills that are required by the criterion task and therefore to be highly correlated with it, that is, to be quite valid One of the more valid tests used in the selection and classification of aircraft pilots by the Army Air Forces is the Complex Co-ordination Test, a "miniature" (life-size but simplified) stick and rudder test which, with its airplane controls and rows of red and green lights, appeals to aspirants to pilot training and involves some of the same ability to co-ordinate arm and foot pressures with each other and with visual stimuli which are involved in actually controlling the plane in flight That its validity is not greater than it is (about 40 with pass-fail in pilot training [214]) is due partly to the fact that response to kinesthetic stimuli, that is to the "feel" of the plane through what fliers call the "seat of the pants," is not required by the test, and partly to the fact that many other factors are important in good flying, especially when the criterion is not just actual *flight* but success in *completing flying training*

The advantages of the miniature test suggest some of its disadvantages A test which seems to have a bearing on an activity which is, perhaps for a quite irrelevant reason, repugnant to the examinee will motivate him

in the wrong manner, as did any test in aviation cadet classification testing which seemed to the would-be pilots to have special bearing on the work of a bombardier One may be able to get a more nearly true measure of the examinee's aptitude or interest with a test the significance of which is not so obvious Steinmetz (754) has demonstrated this with Strong's Vocational Interest Blank, which is not a miniature test but which contains a number of items of obvious vocational significance

Another defect, of less immediate practical importance but more important theoretically and therefore ultimately in practice, lies in the miniature test's unknown elements Since it is a small-scale edition of the task, one has no objective way of knowing what psychological factors it measures. This may be very well in selection testing, when the important thing is to get the highest possible validities with the least possible effort, but in testing for vocational counseling it would necessitate another miniature test for each occupation or at least for each family of occupations to be considered in counseling This would require an inordinate amount of test development and actual testing time It is clearly more practical to analyze each occupation or activity into its important component factors, develop relatively independent tests of each factor or aptitude, validate each of these, and weight each test for each occupation according to its importance in that occupation This makes possible testing for a large number of occupations with a relatively small number of tests It is what was done in the Army's aviation cadet classification program, one test being weighted heavily for pilot, moderately for bombardier, and not at all for navigator, whereas another might be weighted heavily for navigator, moderately for bombardier, and slightly for pilot, according to the demonstrated relationship between each test and the criteria of success in each activity as expressed in correlation coefficients and multiple regression equations The same technique is being used by the Occupational Analysis Division of the United States Employment Service in the development of basic test batteries What the abstract aptitude test loses in validity as a single test of one factor, it generally makes up as part of a battery of tests of known aptitudes combined to give an equally good or better prediction of the same criterion Its principal defect lies in its lack of appeal to the less intelligent examinee, who is not challenged by an abstract task which has no meaning for him and who, if motivated in the right direction, is challenged by a test which resembles an everyday activity For an excellent statement of the case for factorially pure tests, see Guilford (317)

Much of what has been written about miniature and abstract trait tests applies also to *performance* and *paper-and-pencil* tests A performance test is one involving doing something with materials or apparatus, whereas a paper-and-pencil test requires only marking responses to written or perhaps pictorial questions on a sheet of paper The former may be abstract and the latter miniature type, as in the case of the Minnesota Spatial Relations Test and the O'Rourke Mechanical Aptitude Test In the Minnesota Test the examinee places pieces of wood cut in the form of circles, cresent moons, oblongs, and various other shapes in the appropriate holes cut in a board, the assembly has no meaning, other than that of matching different shapes and sizes of objects and holes In the O'Rourke test the subject marks blank spaces to indicate which mechanical objects, tools, etc, are used together or for specific purposes, the task has meaning, in that the objects and processes are taken from real life, are more or less familiar, and serve important practical purposes But in general performance tests have the advantage of being more concrete and therefore seeming to be more meaningful to most people Thus the Minnesota Spatial Relations Test, the real formboard, appeals to some examinees who rebel at the "unreality" of the Revised Minnesota Paper Formboard, a similiar although not identical task in paper-and-pencil form The reason for this is suggested by the relationship between the two tests, expressed by a correlation coefficient of 59 obtained by the writer in an unpublished study of 100 NYA youths, and by the correlations of the two tests with measures of academic aptitude, the formboard having a correlation with the Otis S A Test of 25 and the paper formboard having one of 43 in the same study It appears that the paper-and-pencil test requires more abstract mental ability than the performance test, probably because all spatial manipulations in the former must be made mentally, in abstract form, rather than with actual materials, in concrete form Paper-and-pencil tests are, because of the ease of group administration, cheaper than performance tests once they have been developed, and are often cheaper to develop because of the materials involved

Another dichotomy is that of *tests* as contrasted with *inventories* These concepts are probably familiar enough to need little comment, other than the statement that the former are objective in that they require no judgments of self by the examinee, while the latter are subjective in that they ask the subject to judge or describe his interests, traits, or abilities

It is frequently stated that tests have right or wrong answers, whereas inventories have no right or wrong answers, what is right or wrong in the latter depending on what is true of the examinee This definition is correct when applied to tests of intelligence and to inventories of personality or interests, but it is not correct when applied to tests of personality such as the Rorschach and Murray tests, which are objective and not self-descriptive but which have no right or wrong answers It is also not true of a type of personality test developed in military aviation, which is objective but in which the correct answer is sometimes the wrong one and a wrong answer is sometimes the "right" one, right, that is, for one who is likely to succeed in certain types of occupations Tests have the advantage of being less affected by the desire to make a good impression and by lack of insight than inventories, but are sometimes more expensive in administration and scoring than inventories This is especially true in the field of personality and interest, although the developments in military aviation testing mentioned above, and some comparable civilian work, suggest that this may soon cease to be true in the field of interests (see pages 476 ff)

A fourth dichotomy into which tests may be classified is that of *speed* tests as opposed to *power* tests, illustrated in the intelligence field by the Otis and the CAVD Tests The relative importance which should be attached to each of these has long been a subject of debate in psychological testing, and also fortunately of research Baxter (52,53), for example, has shown that the Otis Self-Administering Test of Mental Ability, administered as a speed test, is a good measure of what will be done by the same subjects when the test is administered as a power test Tinker (847) analyzed the revised Minnesota Paper Formboard as a measure of speed, power, and level, and found that the first two are highly correlated Both of these studies were made with college students, if they had been conducted with older subjects the results might have been different, as Lorge (482) has shown that older persons do as well as younger subjects on power tests, but are handicapped on speed tests The advantages of briefer and uniform timing suggest the use of speed tests with younger persons, and power tests with persons in their forties or above Perhaps the one valid reason for using power rather than speed tests with younger subjects is that speed in the paper and pencil situation is not identical with speed in the life situation, but this is still a matter of supposition which has not been put to experimental proof

Finally, there is the dichotomy of *individual* versus *group* tests, illustrated in intelligence testing by the Wechsler-Bellevue and the Otis or American Council Tests In the former one has the advantage of being able to observe individual reactions and to adapt directions to the intent of the test, rather than having to follow their letter because a modification which would be fairer to one might handicap another subject in the group In the latter social stimulation, competition, the safety of numbers, the group example, and externally standardized conditions facilitate good results

In vocational testing the optimum conditions vary with the circumstances and with the personality of the examinee Sometimes it is better to test an individual alone, whether with group or individual tests, sometimes it helps to have him take tests as one of a group In school situations the latter is more often the case, in a consultation service for adults the former is frequently better policy, although small groups are acceptable In vocational selection, candidates actively seeking employment are probably just as well tested in groups, except in the case of applicants for higher level jobs who may feel that they deserve individual treatment In military selection and classification group testing probably gets better results whether tests are taken voluntarily or by prescription As will be seen in the next chapter, group testing requires either small groups or a large group divided into sections each with its own proctor who can observe it, supervise it, and give attention to special cases

In view of the frequent psychological (and financial) superiority of group testing it is desirable for vocational tests to be suitable for use in groups, they can just as easily be administered individually when that is preferable A limited number of tests can be administered only on an individual basis or to groups of four to six examinees, these should of course be used when they add to the efficiency of the battery and improve the quality of the diagnosis There are no inherent qualities in either group or individual tests which make one type generally better than the other, they must, rather, be considered on the basis of their own validity and of the situation in which testing is to be done Sometimes a test can be so constructed as to be either a group or an individual test in every sense of the term the two forms of the Minnesota Multiphasic Personality Test are an example It would be helpful to see some good studies of the validity of the two forms, the opinion of the test's authors (356) is that when it is administered as an individual test the subject considers each item (printed on a separate card) more carefully and responds more

truthfully than when it is administered in the group form (printed in booklets) and one item closely follows another

The next chapter deals briefly but systematically with methods and problems of test administration, both individual and group, from the point of view of the user of vocational tests, leading up to the chapters which treat specific aptitudes and tests in considerable detail

CHAPTER V

TEST ADMINISTRATION AND SCORING

A PSYCHOLOGICAL test is a measuring instrument The reason for using measuring instruments rather than guesses or judgments based on unaided observation is that psychological tests, like rulers, micrometers, calipers, and scales, are more accurate than the naked eye Since the fundamental reason for resorting to psychological tests is the accuracy of which they are capable, it should go without saying that the user of tests should take pains to give them according to the directions and to do everything possible that will assure accurate results And yet, in everyday practice, one observes countless careless errors in the use of tests, some of them probably not important, but others of vital importance A few such are described in the following paragraphs

The Minnesota Spatial Relations Test was originally designed and standardized as a black formboard, the small pieces which fit into the varied shaped holes also being painted black on top (see p 285 ff) Although none of the original publications dealing with it so state, it was administered in the validation studies with the subject standing (personal letter from Professor Donald G Paterson, dated August 14, 1946) And yet the copies of the test, supplied by one well-known manufacturer and publisher of test materials, are painted black with *green* inserts which probably change the visual problem involved and perhaps make it easier, and the test is administered in some consultation services with the subject standing, in others with him sitting, and in still others either way, according to the client's preference! The writer and a colleague (Charles N Morris) made a study of the effect of taking the tests in these two different ways, with the somewhat inconclusive finding that the presumably improved perspective which is associated with standing above the formboard tends to result in better scores The problem of color has not been investigated, but it seems likely that, contrary to widespread custom, the available norms can legitimately be used neither with the green inserts nor when the examinee is seated Wilson and Carpenter (933) have shown

that the norms for the Crawford Spatial Relations Test, based on the original aluminum form, are not applicable to the marketed wooden form

A consultation service psychometrist was giving the American Council on Education Psychological Examination to a client whose other test results seemed conflicting There was some informal conversation first, after which the examiner rather casually read the directions and proceeded with the test While working on the first timed part the client was puzzled and asked a question in the same informal way in which the proceedings had been conducted from the start The psychometrist answered the question in some detail, then, realizing that some time had been used in which the examinee should have been working on the test, allowed an extra minute for that part As a result of both of these errors, the score could be considered only a crude measure and the client's intellectual status was still not definitely known

In scoring a test used in a large-scale testing program, a clerk failed to invert the scores in order to change the high time scores (score=number of seconds) to low rank scores This error resulted in giving high standing to those who had the least aptitude, and low standing to those who had the most Fortunately, the error was caught in a routine audit in time to prepare a new set of reports, had it not been, time and money, not to mention human energies, would have been wasted when many of the poorer risks failed to make good in an assignment to which they should never have been sent

Perhaps the cause of errors such as the above lies in the very simplicity of the directions for giving and scoring most tests The novice's reaction is that anyone can give most tests, if he knows how to read, and it is true that they are written out so that one should know exactly what to do But their simplicity is deceptive, and errors are frequently made both in following the directions too slavishly when they are poorly written, inappropriate to the situation, or not sufficiently precise, and in departing from the directions when there is no need to do so, in ways not true to their intent For this reason it is necessary to devote some space to the methods and problems of test administration, even when a background of knowledge of the field of measurement is taken for granted

Arrangements for Test Administration

Freedom from distractions is one of the first considerations in providing space for test administration If the examinees are to be free to concen-

trate on their work they must, obviously, not be disturbed by people, incidents, noises, or views which attract their attention away from the tests This seems very simple, until one attempts to define distraction Studies of the effects of noise on work have shown, for example, that typists are able to do as much work, with as high degree of accuracy, under noisy conditions as under quiet conditions, although more strain results in the former (901 506–511) In the group testing of aviation cadets the presence of low-flying planes overhead, where they could not be seen, appeared to have no distracting effect on cadets actually taking tests, although if an especially low-flying plane could be seen it attracted some eyes Super, Braasch, and Shay (803) found that "normal" distractions had no effect on test scores in an experiment with graduate students Apparently a great deal depends upon how much the examinee wants to exclude the distracting factor from his attention if he is well motivated, incidental noises will not bother him, whereas if he is not interested in doing well on the tests he will seize upon the slightest excuse for attending to other matters As one cannot always take good motivation and good work habits for granted, the examiner must take what precautions he can to insure freedom from distractions This means that he should have the use of a room through which there is no passage and to which no one needs to have access during testing, a room without disturbing views of passersby in the corridor or outside the windows, a room not affected by noise in adjacent rooms, corridors, or play space, a room in which the temperature is normal and constant

Good working space for the individual examinee is a second consideration, whether testing is on an individual or group basis In the former this means a table for the examinee, so placed that the examiner can sit opposite him, and a second table so placed that the examiner can reach and manipulate the test materials easily and inconspicuously In group testing, good working space consists of a flat top large enough for the examinee to be able to rest his elbows without touching the persons next to him and to spread out his papers without exposing them to the eyes of his neighbors, this may be made somewhat smaller on especially constructed testing tables by building upright partitions about ten inches high to shut off the view of the neighbor's work Tablet arm chairs such as are used in many college lecture rooms are not desirable for timed tests, especially if separate answer sheets are used, as Traxler (867) has demonstrated These two considerations of sufficiency of space and privacy of work are disregarded with surprising frequency One can some-

times make the best of crowded conditions by using more than one form of the same test

Advance preparation of materials insures having everything needed during the testing (scratch paper for some parts of some tests is frequently forgotten in large-scale group testing), cuts down the time needed for test administration, and results in better morale among examinees In group testing this involves preparing a list of items needed, from pencils to test blanks, and of the quantity to be provided, sorting the materials according to type and sequence in which they are to be used, and counting them out according to the number of subjects to be seated in each row and the number of rows in the room This last step saves a great deal of time and confusion in handing out materials, and prevents the pocketing of excess copies of confidential test booklets In individual testing the steps are essentially the same, but more attention is focused on placing the materials on the examiner's table for maximum availability during testing

Good proctoring is a prerequisite of good testing which results only from securing the assistance of enough proctors and seeing to it that they understand their work The experience of persons working on large-scale testing programs with both students and military personnel has led to recognition of the fact that, when large numbers are being tested, there must be one proctor or testing assistant for every 20 or 25 examinees, if fewer proctors are provided, supervision is likely to be inadequate The functions of the proctors are to distribute test materials, collect them after use, provide sharp pencils when needed, be alert for problems arising from inadequacy of materials (e g , a blank page where there should be printing), from insufficient grasp of directions the understanding of which is assumed once directions have been given (e g , marking answers on the booklet instead of on the separate answer sheet when provided), from "bugs" or defects in the test or test directions which should be recorded for the future improvement of the test, and from abnormal personality traits or poor motivation on the part of examinees Proctors work most effectively when they have not only studied the tests and test directions, but also taken the tests and administered them In large-scale testing operations which have considerable continuity the establishment of training programs to provide for these experiences is not uncommon, in other testing programs the administrator should make the best provisions for familiarizing the proctors with the tests and with the problems which may be encountered in administering them that the situ-

ation permits Testing assistants easily get the feeling that their function is a routine one with neither responsibility nor glory Everything the administrator can do to make them aware of the responsibility they carry and thus insure their careful attention to their work is therefore worthy of consideration

The duration of testing, assuming that more than one or two tests are to be given, depends on the maturity and motivation of those taking the tests In testing for the vocational guidance of high school juniors and seniors and of college freshmen the writer has found two days of testing and filling out records, consisting of three hours each morning and two hours after lunch, quite acceptable to the students The College Entrance Examination Board (23) found that fatigue played no part in six hours of testing When motivation is not so strong and co-operation not so good, periods of two hours may be all that are wise, and there may have to be fewer periods For example, when testing returned combat fliers in an Army Air Forces Redistribution Station it was found that two three hour test sessions, both on the same day, were feasible, but the examiners and proctors needed special skill at times in the handling of recalcitrant officers and men who balked at the length of the testing period Even in this military situation tact went further than authority, and one of the best examiners with poorly motivated returnees was a civilian woman psychologist who knew how both to jest with and to mother belligerent gunners and bombardiers Making clear to examinees why they are taking the tests and how the results will affect them (discussed in a subsequent paragraph) and letting them know at the start just how long the test sessions will last are two essentials to the winning of co-operation in test administration If the examinee wants to understand himself, wants to get a job, or wants to help others like himself (the desire to help other fliers who were *going* to combat motivated many *returnees* in the completing of research questionnaires), he can put in more than a full day of taking tests

Provisions for the recording of the proceedings should also be made ahead of time Decisions should be made as to the type of records to be kept, and appropriate forms provided The times at which tests are begun and stopped should be recorded, as they are occasionally needed later when checks are being made on accuracy of timing Problems arising during testing should be noted, for their value in interpreting the behavior of individuals or the significance of the test results Examiner and proctors both have a part in this work

Testing individuals in groups is a practice frequently made necessary by lack of space and personnel even in consultation services and business enterprises where schedules and needs vary from one person to the next When there are a number of persons to be tested with different tests, and fewer examiners and rooms than there are batteries to be administered (a chronic condition in guidance centers and personnel offices), there is no alternative The space must then be arranged so that individuals or small groups can be sufficiently isolated from others in the same room so that they can work undisturbed by directions not intended for them, stocks of materials must be kept in such a way as to make them readily available to all examiners as needed, and each examiner must develop skill in using several stop watches or chronometers and in shifting from individual to individual as timing requires Space must, of course, permit easy circulation of examiners and of entering and departing examinees

The Preliminaries to Testing

The checking of all arrangements discussed in the preceding section is naturally the first preliminary prior to the starting of testing, in order to be sure that everything necessary is ready for things to go as planned Test administration seems so very simple to the average examinee that its smooth progress is important to rapport

The introductory or motivating talk follows immediately after the arrival and seating of the examinees In prior informal contacts examinees often ask questions of examiners or proctors, thereby demonstrating the widespread need for orientation to that which is to take place, even when testing is voluntary and sought after The knowledge that he is about to put himself to a test or proof makes the examinee somewhat insecure and self-conscious, so that he wants reassurance or feels the need to be somewhat aggressive and belligerent The examiner or proctor, knowing this, can accept his remarks in a calm and friendly way, stating perhaps that something will be said about the nature of the tests before they are started The motivating talk should be brief and to the point Its objective is to set the stage for effective testing by giving the subject some idea of what he is going to do and how long it will take him, and to make him want to portray himself accurately on the tests by relating the taking of the tests to his goals In vocational counseling the goal is self-understanding and better adjustment to the world of work, in vocational selection it is the obtaining of a job in which he will find success and satisfaction These themes can be elaborated upon in ways appropriate to

the age and occupational level of the examinee, but it is well to be sure that the goals are real to those being tested and that the language used in discussing them is appropriate both to the examiner and to the examinee

The Sequence of Tests

In formal testing programs the nature of the tests which need to be given to a particular individual or group determines to some extent the sequence of tests which can be administered Within these limits, however, it is desirable to arrange the order in the way which is likely to interest the examinee most and to get the maximum co-operation from him As a rule, the following principles have been found effective in arranging the sequence of tests

The first test in the series should be something of a *buffer*, one on which the examinee can warm up, get some self-assurance, and develop some interest For this reason it should not be too hard, should be relatively impersonal (i e , neither an intelligence nor a personality test) and objective, and should have "face validity" or seem pertinent to the reason for taking tests (i e , it should, in the case of pilot selection, look like a test that has something to do with flying an airplane)

Next should come a test or tests with long and difficult directions, *difficult content*, or other characteristics which make desirable an alert mind, ability to concentrate, and willingness to apply oneself Tests of this type might come after one or two of the first type, depending on the number and length of those in each category, or they might alternate

Tests which the examiner prefers not to have remembered in detail, if there are such, should come late in the sequence but not at the very end Personality inventories which contain *touchy items* or which might be joked about afterwards are in this category If taken after the difficult tests and before all the other tests have been given they provide some variety and relaxation when it is needed and are likely to be half-forgotten by the time testing is finished

The last test should be relatively *short* and *pleasant*, to help the examinee leave with a good taste in his mouth If a group is being tested together, it is often desirable to let the test be a speeded test so that all may stop and leave at the same time, as having some leave while others are working tends to make the latter finish hurriedly or carelessly, and keeping those who have finished for more than a few minutes is difficult because of restless eagerness to leave When testing individuals in a group

with different tests, untimed tests or inventories may be satisfactory to finish with, the individual can be left more or less to his own devices and others can be given attention

Informal testing characterizes much counseling work carried on entirely by one counselor utilizing interviewing and other techniques (113) Then nothing approaching a "test battery" is administered, but certain tests are used as questions come up on which it is believed they will throw light In such testing the question of sequence is settled by the factors making testing seem desirable the question to be answered provides the motivation for the test being used The problem is then entirely one of selecting an appropriate test rather than one of arranging the tests in the best possible order Bases for test selection are made clear in later chapters of this book

Following Directions in Testing

It has already been pointed out that the very ease with which tests are administered breeds errors Group test administration is likely to be thought of as requiring less skill than other testing operations, group test proctors in aviation cadet classification testing referred to themselves colloquially and collectively as the "bunion brigade" Unless examiners and proctors are aware of the ease with which errors are made and are challenged by the need for care, they are likely soon to be guilty of unknowingly modifying the introduction to testing in such a way as to change the examinees' motivation for better or for worse, of changing directions in ways which give them either more or less help than they should have in taking the test, of answering questions which give then an unfair advantage in comparison with the groups on whom norms were established; and even of allowing too much or too little time in which to take the tests

The writers of *motivating talks* and of *test directions* intend to convey ideas to the examinees which will motivate them in certain ways and lead them to work according to certain methods *If*, therefore, the examiner understood exactly what the test constructor intended to convey, and *if* he were able to express that idea just as clearly as the test author in words of his own, there would be no reason why he should not rephrase the directions to suit himself and vary his statements from time to time Unfortunately, however, experience has demonstrated time and again that while the modifications made in test directions may be just as clear to the examiner as were the originals, they are rarely if ever as clear to the

examinee The reason for this is obvious enough the directions supplied with a well-constructed test have been tried out a number of times on subjects like those for whom the test was developed before it was finally published, and each time they were rewritten somewhat and improved after criticism by examiners and examinees in order to make sure that the intended meaning and the understood meaning were identical Obviously, the directions more or less casually phrased and even more casually tested by the user of a test are not likely to be as clearly and as uniformly understood as those that are printed with the test Only a highly skilled examiner who knows both his test and his subjects well should allow himself the privilege of improvising or modifying directions At the same time all examiners need to scrutinize the printed directions carefully to be sure that they are well drafted If they are not suitable for the group in question the suitability of the test itself may be open to question, if the test is suitable with different directions, the norms may no longer be applicable These matters are subject to empirical check, and if judged important enough the answers may be found by experimental methods It is good practice for examiners to have a manual or loose-leaf notebook of test directions, and to know these well in order to facilitate reading them while administrating tests

Examinees' questions need to be viewed by the examiner as possible requests for changes in the test directions If the information asked for was supposed to be conveyed by the directions, and if understanding of the directions was supposed to be achieved before beginning the test (rather than being a part of the test), the examiner should answer the questions promptly and concisely If, on the other hand, answering the question would give the examinee an understanding of the test or information which the directions were not intended to convey, to do so would be to make his score meaningless, or at least impossible of comparison with those of others who took the test and on whom the norms were based In such a case the best answer is "That's for you to decide" or some equivalent which makes it clear that the examinee must find the solution himself It should be stressed that the number of questions asked, and their legitimacy, depends to a very considerable extent upon the manner of the examiner If he gives directions in too businesslike and cold a manner questions which should be asked will not be voiced, if he is too informal and friendly too many unfair questions will come up, but if he gives directions clearly and pleasantly he will meet with an optimum number of questions concerning matters included in the directions (few

but all necessary questions) and a minimum of questions of types which he should not answer

This leads to the topic of the examiner's *voice and attitude*, both of which have considerable effect on the attitudes of examinees and therefore on the validity of the tests which they take An examiner whose clear, confident, and friendly voice and interested alert manner are noted by the examinees gives them the feeling that the tests are important, interesting, and worth taking seriously, one who is lackadaisical in manner, fearful in front of a group, or careless in his speech is not likely to create in his subjects attitudes which make for serious application and genuine co-operation When proctors assist in test administration, the manner in which they walk the aisles and watch examinees or stand idly by with their minds obviously far away is equally important

The need for *accuracy of timing* has already been mentioned In administering tests of manual dexterity or other aptitudes best measured by apparatus tests this necessitates a stop watch with its easily controlled second hand Most paper and pencil tests, however, can be timed with sufficient accuracy by means of the second hand of an ordinary watch if the hand is long enough A watch with a sweep-second hand is even better, although still not as easily used as a stop watch because the examiner must watch the second hand enough to count the number of times it goes around This is best done by tabulating on a pad If a stop watch is available, the freedom to spend more time watching examinees and less time looking at the second hand is desirable As stop watches are sometimes erratic, it is advisable to check their operation before testing, and to note the starting time on one's wrist watch or on a clock (or to have a proctor time with a second stop watch) in order to be sure that watch trouble does not prevent accurate timing of a test that is actually under way Finally, when testing large groups it is good practice to instruct examinees to put their pencils down and lean back in their chairs at the word "Stop," thus making it easy for examiner and proctors to insure the respecting of time limits on strictly timed paper and pencil tests

Observing the Behavior of Examinees

Careful observation of the manner and attitude of the examinees has long been standard practice in clinical testing, and has been carried over into vocational testing by those with clinical training Baumgarten published a list of types of behavior which should be looked for by the user of vocational tests (51), and Bingham translated and converted this

into an Examiner's Checklist (94 229–235) An abbreviated and somewhat modified form of this checklist is included here because of its value in suggesting types of behavior which may be worth noting and the possible

EXAMINER'S CHECKLIST

Examiner _____ Subject _____

Date _____ Test _____

BEHAVIOR	*INTERPRETATION*
I *PRELIMINARIES* A Attention to Examiner 1 Attentive_____ 2 Looks around_____	
B Questions 1 Yes_____ 2 No_____	
C Speed of Approach to Test 1 Rapid_____ 2 Slow_____ 3 Hesitating_____	
D Seriousness of Attitude 1 Serious_____ 2 Playful_____ 3 Zealous_____	
E Confidence in Approach Disparages 1 Disparages task_____ 2 Enthusiastic_____ 3 Self_____	
F Judgment of the Task 1 Vocal_____ 2 Gestures_____	
II *EXECUTION* A Starting 1 Deliberation a Yes_____ b No_____	
2 False Starts a Yes_____ b No _____ a Perseverates_____ b' Changes_____	
B At Work 1 Direction of Attention a To task_____ b Away _____	
2 Degree of Attention a Concentrated_____ b Distracted_____	
3 Expression of Feelings a Yes_____ b No_____	
4 Body Movements a Co-ordinated_____ b Not co-ordinated_____	
5 Hand Movements Yes_____ Yes_____ Yes_____ a Appropriate b Sure c Quick No _____ No _____ No _____	
6 Work Tempo a Quick_____ b Slow _____	
7 System a Yes_____ b No_____	
8 Regularity 1)Crescendo _____ a Yes_____ b No_____ 2)Diminuendo_____ 3)Alternately _____	
9 Care and Neatness a Careful_____ b Sloppy_____	
C Frustration Tolerance 1 Asks no help _____ a Indifferent _____ b Gives up _____ c Solves problem_____	
2 Asks help_____ a Once_____ b Repeatedly_____	
3 Receives help _____ a Indifferently_____ d Critically _____ b Happily _____ e Trustfully _____ c Gratefully _____ f With offense_____	

EXAMINER'S CHECKLIST CONT'D

D Obedience to Instructions 1 Exact_____ 2 With Deviations_____	
III *ATTITUDE TOWARD PERFORMANCE* A Notices Mistakes_____ a In process_____ b At end_____ c Sporadically_____	
B Mistakes Unnoticed_____	
C Shows Feeling_____ a Pleasure_____ b Vexation_____ c Not clear_____	
IV *CONDUCT AFTER TEST* A Silent and Watchful_____	
B Announces Result_____	
C Asks Evaluation_____	
D Expresses Feelings_____ a Satisfaction_____ b Vexation_____	
F Leaves Materials a In order_____ b In disorder_____	

significance of such behavior, and because it is useful in training psychometrists, counselors, and personnel workers to get more than a test score from the administration of a test In actual practice, however, such elaborate forms are rarely used instead, the examiner who has learned to observe behavior in testing simply makes note of anything which he believes may be significant and includes it in his test report Beginners do well to use a form such as this for some time, in order to learn what to look for and to get the habit of noting it, once the habit has been acquired the simpler method can effectively be adopted instead Examples of notations of behavior in taking tests and of their use in interpreting test scores are given in Chapters 21 to 23, in which methods of reporting test results are discussed in some detail and the content of test reports is illustrated

One word of caution should be said at this point Some clinicians delight in telling how much is learned about a subject from the way in which he attacks a problem, from his procedure in putting together a set of Wiggly Blocks or from his persistence in working on a difficult mechanical problem These symptoms are extremely interesting, and it is easy to be carried away by the tendency to build an ambitious account of a personality upon them They are, however, minute segments of behavior observed in a limited situation, and there is no real evidence that the behavior so manifested is typical of behavior in other situations The possible insights which may be gained from watching a person solve arithmetic problems while taking a standard test should not be missed, but it should be remembered that it is the *score* which has been proved reliable and which is known to be related to behavior in other situations,

not the *method of approach* or the *reaction to frustration* At the same time, a knowledge of these latter helps one to understand how and why the obtained score was obtained, and provides data which may, with many other items from other situations, help in the construction of a picture of the counselee's personality

Condition of the Examinee

Those who take objective tests sometimes claim that they are too nervous at the time of testing to do themselves justice, or that they were not in good health at the time and were therefore handicapped In certain extreme cases these claims are no doubt warranted, as for example in that of a married man who took a test the morning after a violent quarrel with his wife and a subsequent resort to alcohol he was in the second decile of the comparison group in that testing, but on a retake three months later, after a divorce and when clear-headed, he was in the highest decile Despite these occasional rather obvious and verified cases, there is a good deal of skepticism concerning most such claims Oddly enough, there has been very little research on these problems

The influence of tension on the intelligence test scores of children was investigated by Yager (947) with a group of forty boys from ten to twelve years old They were first tested under normal conditions, then under tension presumably produced by threats and evidenced by physiological changes Thirty of the boys made better scores, but ten showed losses The tendency to improve or to break down under tension was related to emotional stability This experiment appears to confirm the belief that only a few persons, and those the neurotically inclined, suffer from the tension-creating conditions of testing

The effects of health were investigated by British army psychologists in a study referred to by Vernon (897) Standard selection tests were taken a second time by women recruits, and differences were related to menstrual phase The effects of menstrual cycle on test scores were found to be negligible Another group of over 1000 were asked at test and retest whether or not they felt able to do themselves justice, less than four percent claimed not to be able to do themselves justice, but their scores were not significantly different from those of the others Those suffering from colds showed a slight, but not significant, drop in scores

Another study which may have some bearing on this problem is a report by Glick (292) that freshmen who took the college intelligence tests during the New England hurricane of 1938 made scores 20 percent

higher than those of other years When subsequently retested, they were shown to be a normal group Glick suggests that their "hurricane intelligence" may have been the result of stimulating effects of ozone in the air at the time of the hurricane

These studies, like those of the effect of distractions on test scores, suggest that the minor illnesses which do not confine one to bed can be dismissed as having no appreciable effect on test scores, but that the more serious impairments are sufficient justification for questioning a test score

Scoring Tests

The methods of scoring the tests which are widely used in vocational guidance and selection are objective and generally quite simple The tests in which scoring involves judgment and training on the part of the examiner are used almost exclusively in clinical work, exceptions to this statement are the clinically interpreted Wechsler-Bellevue Test of Intelligence, sometimes used as a special check in cases of adults who may be verbally handicapped, and the Rorschach Psychodiagnostic, which is occasionally used in connection with executive selection Both of these require extended training of a type which is given in special courses, and are the subject of special books (914, 56, 57, 108, 433) Most of the tests which are widely used and which are discussed in this book are scored by means of stencils or keys which can be used by a clerk or by clerically operated scoring machine, the others have simple time scores For this reason only two points need to be made concerning the scoring of vocational tests

The first of these is, again, familiarity with the directions Persons scoring tests must first be sure that they understand the procedure A routine can then be established that fits the immediate situation If hand scoring is in order, clear and durable keys or stencils should be made, and scores should be systematically calculated and entered on record blanks If machine scoring is done (it should be in any large-scale operations) this work will either be performed by a commercial scoring organization or by an especially trained scorer who is competent to set up procedures

The second point has to do with checking Even the best of scorers make errors, as illustrated at the beginning of this chapter For this reason all scores should be checked by another person, at all stages, if hand scoring is utilized If machine scoring is used, all manual steps should be checked If an accurate instrument is worth using, it is worth insuring its accurate use

CHAPTER VI

INTELLIGENCE

Nature and Role

INTELLIGENCE has frequently been defined as the ability to adjust to the environment or to learn from experience As Garrett (281) has pointed out, this definition is too broad to be very helpful in practical work One might therefore resort to an operational definition, and say that intelligence is the ability to succeed in school or college such a definition would be justified by the fact that the criterion used in standardizing intelligence tests has generally been one of school placement and progress This line of thought is illustrated by the tendency of many school and college officers to talk in terms of scholastic aptitude and scholastic aptitude tests, thereby implicitly limiting the application of such tests to the situations in which they have been proved valid and dodging the issue of their role in other types of situations

An equally operational, but more psychological and therefore more generally applicable, definition is suggested by Garrett in the paper referred to above "Intelligence ," he states, "includes at least the abilities demanded in the solution of problems which require the comprehension and use of symbols" This definition is operational in that it is based on an analysis of the task involved in solving the problems presented by an intelligence test It is broader than some test-based definitions because it applies not only to the tasks presented by the test, but also to the tasks presented by the school or college courses, success in which it is designed to predict It is broader even than this, because it allows for the value of such tests in predicting success in certain types of occupations, namely those in which job analysis shows that it is necessary to comprehend and use symbols And it has the additional advantage of taking into account the important work of the past ten or fifteen years which demonstrates that intelligence is not one aptitude but a constellation of aptitudes As these components of intelligence apparently vary in importance in different occupations according to the type of symbol

most frequently used in that occupation, this advantage is of great practical significance

Two closely related questions normally come up for discussion at this stage those of the innateness and the constancy of intelligence During the 1930's they were the subject of much debate and disagreement among psychologists, an excellent overview of which is provided by the 39th Yearbook of the National Society for the Study of Education (920), reference should also be made to a paper by Stoddard (760) expounding the environmentalist point of view, and to papers by McNemar (501), Thorndike (832), and Wellman et al (917, 918), in which detailed questions of the methods and results of nature-nurture studies are examined at length The topic is much too complex for treatment in a handbook on vocational testing The reader who has not studied sources such as those referred to, or who has not sufficient time to do so, must rest content with the general conclusion reached by this writer This is that whereas both nature and nurture play a part in the development of intelligence, mental ability as indicated by the intelligence quotient is relatively constant from the time a child enters elementary school until late adulthood It is true that the obtained I Q will vary some after the age of six, but this is generally more a function of the tests, which are often not strictly comparable at different age levels and which are in any case subject to errors of measurement, than of the individual Some changes which are too great to be explained by these causes are the result of emotional conditions which invalidate the score of one test, or of organic changes resulting from disease or injury That there are other changes, not explained by any of these factors and attributable to changes in the environment which modify the functioning intelligence, has not been demonstrated to the satisfaction of all competent judges with persons of elementary school age or older

Intelligence and Educational Success

The role played by intelligence in educational achievement has been frequently studied Comprehensive reviews of the research are available in Pintner (604 Ch 10–12) and in Strang (766 72–92) Our attention will be focused on certain points, an understanding of which is needed in the use of intelligence tests in educational and vocational guidance, and on some data illustrating those points

Different *curricula* have been found to require or to attract different degrees of intelligence, whether at the high school or at the college level

In general, students in scientific and liberal arts courses have the highest intelligence test scores, with those in commercial subjects coming next and trade courses last In one nation-wide study (417) the median I Q of high school boys in different courses was as follows

TABLE 3

MEDIAN I Q 'S OF BOYS IN HIGH SCHOOL COURSES

Course	*Md I Q*
College Preparatory (Technical Schools)	114
Scientific (General Schools)	108
Academic	106
Commercial	104
Trade	92

The exact figures vary from one community to another and from time to time It is therefore necessary to have local norms in actual counseling, in fact, not only the trends, as indicated by averages, are necessary, but even more needed are minimum critical scores which show what score a student should make in order to be a good risk in each type of training The importance of local norms is further illustrated by the fact that in some cities, Buffalo for example, there are trade schools which offer such attractive training that entrance is quite competitive, whereas some of the general high schools attract students of less ability who, for cultural reasons such as the prestige of academic training, want the traditional education It should be remembered, too, that if general intelligence were broken down into its component factors, the group which ranked highest on one might well rank lower on another

Differences in the intelligence scores of students in different *institutions* have been found which, like curricular differences, are in line with popular expectation Some of these can be expressed in generalizations liberal arts college students tend to be intellectually superior to teachers college students, those in small rural colleges tend to be inferior to those in large urban universities, and those in highly endowed private institutions tend to be more able than those in state universities (at least when freshmen classes are compared) or in denominational colleges The documentation for these statements is provided by the periodic analyses of the results of nation-wide testing programs such as that of the American Council on Education (840), in which some 350 colleges and universities of all types usually participate After World War I studies made in a number of universities with the Army Alpha Intelligence Test gave re-

sults for a larger group of identified institutions than more recent publications, which generally use code numbers rather than names The data have been collected by Pintner (604 296), converted into Otis I Q equivalents, these reports show that some twenty years ago the median I Q at Yale was 131, at Oberlin 124, at Ohio State 120, at Penn State 117, and at Purdue 115 The overlapping of scores was no doubt considerable, but the ranges and quartiles are not reported

The American Council data referred to above are for entering freshmen, which means that the normal elimination as a result of academic failure has not yet taken place This is especially important at state institutions which are obliged to admit great numbers of high school graduates who subsequently fail to keep up with their classes, and which therefore have freshman attrition rates as high as 50 and 60 percent In colleges using more stringent selection standards the differences in the average intelligence of freshmen and seniors is much smaller Really adequate data on intelligence and college success would, as in the case of curricula, provide minimum critical scores for each college Individual colleges, as will be seen shortly, have such data for their own use The published material, however, is simply in terms of freshman averages and variations In 1938, for example, the 355 colleges using the A C E Psychological Examination (840) reported freshmen medians which, when converted into Otis I Q equivalents, range from 94 to 122, the median college having a median freshman I Q of 108 The interquartile deviations were such that the college with a median freshman I Q of 94 had a freshman class in which one fourth of the students had I Q equivalents of less than 90, and only one fourth exceeded 100

Data for one liberal arts college, Oberlin, have been reported in some detail by Hartson (348, 349), who has set an example which, if followed by other college officials, would be of general benefit in improving the college counseling done in the high schools At Oberlin some students with Otis I Q equivalents of less than 100 manage to graduate, but Hartson found that 65 percent of the entering freshmen who were below 110 failed academically In another college of approximately the same academic but lower social standing, it was found that there were practically no freshmen with I Q's of less than 110, indicating that the latter institution was admitting students on a more selective academic basis Attrition data also showed a higher mortality rate among the lower intelligence levels at the latter college Obviously, the former institution

would be a better choice for a student with an I Q equivalent of about 110 At Franklin and Marshall the mean I Q was 111, but here also 65 percent of those below 110 failed (512)

Despite the relationship between intelligence and educational achievement revealed by data such as the above, the correlation between intelligence tests and *grades* is not especially high The numerous summaries of the subject show that in high school they tend to range from 30 to 80, and in college from 20 to 70, the modal r s being 40 and 50 in the former and between 30 and 50 in the latter The relationship in college seems lower than in high school because the selection procedures in colleges cut down the range of ability in their populations, and this in turn makes the correlation coefficients shrink artificially The relationships are high enough to make them useful in studying groups, but the margin of error when working with individual students is so great as to make considerable caution necessary in test interpretation and to require that the counselor or admissions officer give considerable weight to other indices such as high school marks, family educational achievement (as an indicator of what his intimate social group expects of him), personality adjustment and motivation None of these, taken by itself, is any more valid than the score of a good intelligence test for predicting college marks, but, taken together, they yield a better prediction than any single index (766 123) To cite the Oberlin studies once more, the fact that 65 percent of the freshmen who were admitted with I Q's of less than 110 failed academically is a legitimate reason for questioning the choice of that college with an I Q of 110, on the other hand it should be remembered that 35 percent of such students graduated The counselor must ask himself, and get the student to ask himself, what reasons there are for expecting him to be in one group rather than in the other, and whether or not a less competitive situation might not be more conducive to his fullest all-round growth

The relationship of intelligence test scores to educational achievement has been demonstrated in one other type of study, in which a genetic approach has related intelligence to *amount of education* obtained These studies make it clear that, on the whole, those who are most able obtain the most education Proctor (613) made a follow-up study in 1930 of persons who had been tested while in school in 1917, and found that those who, in 1930, had gone no further than the 9th grade had an average I Q in 1917 of 105, whereas those who had graduated from high school had a mean of 111 and those who went to college had averaged 116 This should

not, of course, be taken as proof that students who have the ability to do college work manage to go to college the Pennsylvania Study (458) demonstrated that the bright students who actually get to college are matched by an equally able but economically less fortunate group who do not obtain that much education What Proctor demonstrated is that those who get more education are, on the whole, more able than the much larger group who obtain less education

Terman's long-term studies of gifted persons, begun in 1922 and reported in two follow-ups (821, 823), provide some more data which demonstrate the importance of intelligence in completing an education Almost all of his group of 1300 children with I Q 's of more than 140 graduated from college (helped, be it said, by the fact that they lived in a state which provides more low-cost higher education than any other for its residents)

The studies mentioned so far have all dealt with the relationship between intelligence and educational achievement, none with the role of the former in *satisfaction* in one's studies It is generally assumed that the placement of a student at the proper educational level, one on which he can compete with his peers without undue strain and on which he will be challenged by the need to exert himself in order to master the subject-matter, results in better adjustment and greater satisfaction on his part The assumption seems reasonable Every experienced teacher can cite instances in its support The literature of clinical psychology abounds in references to cases illustrating it (140, 487) But oddly enough there are no studies involving objective measures and carefully quantified data to prove the validity of the assumption In one investigation Berdie (78) correlated intelligence test scores and measured satisfaction in the study of engineering, finding an r of 02 This is disappointing, but is probably more a defect in the experiment than in the hypothesis the scale used for the measurement of satisfaction may not have been sensitive to what it attempted to assess, or the relationship may be such that it would manifest itself in a study of many curricula without being revealed in a study of one type of curriculum The latter, it will be seen, is true of the relationship between intelligence and success in occupations Although one is justified in generally being skeptical of clinical experience and professional opinion unsupported by experimental evidence, this would seem to be one instance in which it is best, pending the carrying out of adequate objective studies, to accept the evidence of subjectively analyzed experience This would lead one to conclude that students who are placed

in courses which are difficult enough to make them work but not so difficult as to discourage them are most likely to be satisfied with and interested in their studies

Intelligence and Vocational Success

As intelligence has been supposed to affect vocational success in a number of different ways, tests have been correlated with a variety of criteria These include wisdom of vocational choice, success in training, ability to secure a job of a particular type, adjustment in the world of work as shown by placement on the occupational ladder, status in the occupation as indicated by criteria ranging from tenure to earnings, and satisfaction in one's work Each of these will be discussed in the following paragraphs

Vocational Choice In a number of studies (305, 728, 943) the more intelligent individuals have been found to have more appropriate occupational objectives This is what one would expect *on a priori* grounds, not only because the more able should have better insight into their own abilities and into job requirements, but also because, in a society which encourages people to aspire to the higher levels, they have more of the abilities which are required for success in the prestige occupations The factors considered in these studies have usually been limited in number Sparling (728), for example, compared the tested intelligence of the student with the intelligence considered necessary for success in his chosen field on the basis of an analysis of intelligence test data gathered from soldiers in World War I, while Wrenn (943) compared the correspondence between measured and self-estimated interests at different intelligence levels Atomistic as they are in their approach to these problems, the investigations justify one in concluding that the more intelligent are more likely, other things being equal, to make wise vocational choices

Success in Training This topic has been dealt with under the heading of intelligence and educational success, as most formal training is under educational auspices and bears an educational label But, since one cannot succeed in medicine or flying without first succeeding in medical or flying school, success in training is the first step in vocational success It is frequently much easier to obtain criteria of success in training than in the practice or pursuit of the vocation itself For these reasons training success is a commonly used criterion of vocational success, and needs to be mentioned in this section

Securing employment Studies of the relationship between intelligence and ability to secure employment have been made in depression years, as

those are the times when attention is focussed on the problem of what it takes to obtain a job and on the differences between employed and unemployed workers

Few of the youth studies of the 1930's used measures of intelligence, presumably because they were large scale surveys in which accurate testing was impractical In several studies confined to more accessible subjects, however, testing was carried out with what at first appear to be surprising results Dearborn and Rothney (197) analyzed the relationship between tested intelligence and success in securing employment in a large sample of youth who were subjects of the Harvard Growth Study and lived in communities adjacent to Cambridge, Massachusetts They found no relationship Lazarsfeld and Gaudet (457) studied a small but carefully matched sample of youth in Essex County, New Jersey They also reported no relationship between tested intelligence and success in finding employment

In contrast to these and similar studies of young persons stand the investigations dealing with adults in the depression Morton (545) and Paterson and Darley in their summary of the psychological work of the Minnesota Employment Stabilization Research Institute (589) reported that, in a variety of occupational groups in Montreal and in Minneapolis, the early unemployed were less able than those who were released later in the Depression At least in retaining their jobs, then, the more intelligent fare better than the less intelligent This suggests that in employing young people the average business man either does not have access to or does not utilize data revealing the abilities of the employment applicants, but relies instead on other and, as Dearborn and Rothney showed, less relevant indices, whereas the employer who is considering releasing employees does depend more on indices of ability In the case of a worker already in his employ this need not be, and generally is not, an intelligence test, but is simply the employer's judgment of the relative value to the company (efficiency, versatility, etc) of each of the persons in question No such ability data, which frequently correlate with intelligence test scores (see below), are available to the employer of relatively inexperienced youth, although school experience should be such as to provide employers with data of the same type and intelligence tests can be used in selection Personnel men should be able to make considerable improvement in their work by bringing their practices in employing new workers up to the level of their practices in releasing workers when staffs must be cut

Attainment on the Occupational Scale In a culture in which material success and ability to rise to or maintain a high socio-economic level are valued as highly as in ours, the question of the relationship of intelligence to attainment on the occupational scale is one of vital importance If the relationship is close, then the ambitions of many persons are unrealistic and, if not modified by experience, are doomed to disappointment, whereas if the relationship is not close then there is some justification for the widespread encouraging of youth to aspire to the higher levels

The first large-scale studies of this question were made possible by the mass of data accumulated as a result of the use of intelligence tests in the Army of the United States in World War I These were analyzed and published in the *Memoirs of the National Academy of Sciences* (525), and were subsequently reworked by Fryer (276) and by Fryer and Sparling (278) to make them more usable in vocational counseling Similar data for World War II, based on a sample of some 90,000 white men, have been organized in a similar table by Stewart (758), reproduced on pp 96–97 by permission of *Occupations*

A table such as this is useful in ascertaining approximately the occupational level at which an individual is most likely to be able to compete without undue strain and, at the same time, with sufficient challenge to make the work interesting To know that a student with a score of 125 has the general ability to compete with men and women who have been successfully engaged in the lower professional and managerial occupations, but somewhat less than that which characterizes those who have made good in the higher level occupations of the same type, is of value

But the apparent simplicity of the chart is deceptive because it does not bring out the great overlapping of the various occupational intelligence levels A given occupation actually includes within itself a great variety of levels a chemist, for example, may supervise routine tests on the one hand or do highly creative experimental research on the other, or, more commonly, something in between these two extremes This means that there are opportunities in most occupations for some persons at relatively low levels who are not likely, if their mental ability is appropriate to these levels, to rise appreciably in the field, and for others with greater ability who should, other things being equal, rise to higher levels Thus some chemists really belong in the highest occupational level in Table 4, where the majority are placed, but others should be in the second group of occupations Other factors which play a part in occupational success need, of course, to be taken into account, but are not in the

chart lack of motivation may disqualify a person from competing effectively at his appropriate intellectual level, or an unusually effective personality may enable another to compete above that at which he might otherwise be expected to make an optimal adjustment

The overlapping of occupations when classified according to intelligence is well brought out by Stewart, who reports the median and adjacent quartiles for a number of different occupations in the Army sample For example, a man with an AGCT score of 115 might, in so far as *mental ability* is concerned, be a high-average stock-keeper, average general clerk, low-average bookkeeper or below average accountant—all in the clerical field, not to mention an average draftsman or a low-average reporter in other fields Clearly, the extent and nature of the overlapping is so great that, while occupational intelligence levels provide a rough guide, they must be used as that and cannot be applied in a mechanical or arbitrary way

Another limitation to the value of World War II data is imposed by the nature of the sample Some occupations were not adequately represented in the Army The War having been a total war, and Selective Service having operated according to written directives and on the basis of studies by the War Manpower Commission, we know a good deal how the occupations represented in the Army were affected by sampling problems As lawyers had a type of training which was at a premium in neither war industries nor military service during the early years of the war, it seems likely that the drafted lawyers are fairly representative of the young lawyers of that time Psychologists, on the other hand, were at a premium in both military and industrial personnel work, and as the Army commissioned many who were aged thirty or more, and the Navy many who were under thirty, directly from civilian life early in the war, it is probable that the drafted psychologists who held the Ph D degree at the time of being drafted were not really representative of young psychologists in mental ability and *savoir faire* It is to be hoped that a thorough-going study of data obtained during World War II will be made, relating occupational intelligence findings to known policies of Army, Navy and Selective Service Stewart did not do this

A final possible defect in intelligence test data obtained under military auspices which must be mentioned is the fact that the testing conditions are often not optimal Many new draftees were not well oriented to psychological tests, these often resented the tests as so much mumbo-jumbo Others were negativistic in their attitude toward the Service and vented

TAB

OCCUPATIONAL GROUPS WHOSE AGCT MEDIANS LIE IN EACH
Based on 88,907 White Enlisted Men in

−2.5σ	−2.0σ	−1.5σ	−1.0σ	−.5σ	Mean
85.3	89.9	94.5	99.1	103.7	108.3

Teamster
Miner
Farm Worker
Lumberjack

Marine Fireman
Laundry Machine Operator
Laborer
Barber
Shoe Repairman
Jackhammer Operator
Groundman, Telephone, Telegraph, or Power
Section Hand Railway

Tractor Driver
Painter, General
Foundryman
Animation Artist
Hospital Orderly
Baker
Packer, Supplies
Sewing Machine Operator
Truck Driver, Heavy
Painter, Automobile
Hoist Operator
Construction Machine Operator
Horsebreaker
Tailor
Stonemason
Crane Operator
Upholsterer
Cook
Concrete-Mixer Operator
Truck Driver Light
Stationary Fireman
Warehouseman
Gas and Oil Man
Forging-Press Operator
Longshoreman
Well Driller

Welder Electric Arc
Plumber
Switchman, Railway
Machine Operator
Hammersmith
Student, High School, Agricultural
Automotive Mechanic
Blacksmith
Welder, Acetylene
Bricklayer
Blaster or Powderman
Small Craft Operator
Lineman, Power
Packing Case Maker
Carpenter, General
Pipe Fitter
Electric Truck Driver
Highway Maintenance Man
Automobile Serviceman
Rigger
Woodworking Machine Operator
Chauffeur
Motorcyclist
Burner, Acetylene

Not Elsewhere Classified
Machinist's Helper
Foreman, Labor
Locomotive Fireman
Entertainer
Meat Cutter
Student, High School Vocational
Cabinetmaker
Airplane Engine Mechanic
Heat Treater
Fire Fighter
Engineering Aide
Construction Equipment Mechanic
Optician
Packer, High Explosives
Petroleum Storage Technician
Pattern Maker, Wood
Electrician, Automotive
Coppersmith
Ship Fitter
Sheet Metal Worker
Electroplater
Instrument Repairman, Electrical
Steam Fitter
Diesel Mechanic
Carpenter, Ship
Bandsman, Snare Drum
Lithographic Pressman
Electric Motor Repairman
Shop Maintenance Mechanic
Job Pressman
Riveter Pneumatic
Power Shovel Operator
Photographic Technician, Aerial
Brakeman, Railway
Automobile Body Repairman
Tire Rebuilder
Utility Repairman
Boilermaker
Foreman, Automotive Repair Shop
Salvage Man
Structural Steel Worker
Welder, Combination
Welder, Spot
Seaman
Engineman, Operating
Foreman, Construction
Millwright

LE 4

HALF-SIGMA INTERVAL FROM THE MEAN OF ALL THE MEDIANS

Machine Records Survey Taken June 30, 1944

+5σ	+1.0σ	+1.5σ	+2.0σ	+2.5σ
112.9	117.5	122.1	126.7	131.3

Carpenter, Heavy Construction
Dispatcher, Motor Vehicle
Gunsmith
Musician, Instrumental
Tool Maker
Nurse Practical
Photographer, Portrait
Photolithographer
Rodman and Chainman, Surveying
Airplane Fabric and Dope Worker
Multilith or Multigraph
Shipping Clerk
Printer
Steward
Foreman, Warehouse
Bandsman, Cornet or Trumpet
Instrument Repairman Non-electrical
Boring Mill Operator
Projectionist, Motion Picture
Dental Laboratory Technician
Laboratory Technician, V-mail or Microfilm
Foreman, Machine Shop
Stock Clerk
Painter, Sign
Machinist
Photographer, Aerial
Engine Lathe Operator
Parts Clerk Automotive
Cook's Helper
Railway Mechanic, General
Office Machine Serviceman
Student, High School, Commercial
Electrician, Airplane
Student, Manual Arts
Policeman
Sales Clerk
Electrician
Lineman, Telephone and Telegraph
Watch Repairman
Receiving or Shipping Checker
Car Mechanic, Railway
Toolroom Keeper
Refrigeration Mechanic
Cameraman, Motion Picture
Telephone Operator
Hatch Tender
Switchboard Installer, Telephone and Telegraph, Dial
Cashier
Stock Record Clerk
Clerk, General
Radio Repairman
Purchasing Agent
Survey and Instrument Man
Physics Laboratory Assistant
Stock Control Clerk
Manager Production
Boilermaker
Layer Out
Radio Operator
Linotype Operator
Student, Mechanics
Salesman
Athletic Instructor
Store Manager
Installer Repairman, Telephone and Telegraph
Motorcycle Mechanic
Dispatcher Clerk
Crew
Tool Dresser
File Clerk
Embalmer
Brake Inspector, Railway
Airplane and Engine Mechanic
Shop Clerk
Artist
Band Leader
Photographer
Geologist
Airplane Engine Service Mechanic
Cable Splicer Telephone and Telegraph
Surveyor
Student, High School, Academic
Blueprinter or Photostat Operator
Bookkeeper, General
Chief Clerk
Stenographer
Pharmacist
Typist
Draftsman
Chemical Laboratory Assistant
Draftsman, Mechanical
Investigator
Reporter
Tool Designer
Tabulating Machine Operator
Addressing Embossing Machine Operator
Traffic Rate Clerk
Clerk-Typist
Postal Clerk
Bookkeeping Machine Operator
Meat or Dairy Inspector
Photographic Laboratory Technician
Teletype Operator
Student Sociology
Writer
Student, Civil Engineering
Statistical Clerk
Student Chemical Engineering
Teacher
Lawyer
Student, Business or Public Administration
Auditor
Student, Dentistry
Accountant
Student, Mechanical Engineering
Personnel Clerk
Student, Medicine
Chemist
Student, Electrical Engineering

their feelings in not co-operating in the testing—often to their regret when they found, later, that they needed a higher score in order to qualify for officer candidate school (an error many remedied by retaking the test and making qualifying scores) Still others, heeding rumors that men who made high scores were being assigned to a type of training they did not want (for example, to Link trainer instruction when they wanted to be aerial gunners), made low scores in order to avoid it But draftee attitudes were not the only problem Some were created by "efficiency" minded or routine-bound officers who sent men to testing after a night of duty in the kitchens or after they had had only a few hours sleep subsequent to a long trip by troop train But this should not lead to the conclusion that all military testing was conducted under poor conditions or that the results should be entirely disregarded On the contrary, much of it was well done, and many, probably most, of those who took the tests tried to do their best It is easy for a few dramatic cases to create a false impression in such a situation

The trends revealed by military studies have been confirmed not only in studies abroad by Cattell (151) and Awaji (34), but also in civilian studies made in this country by Scott and Clothier (685) and Pond (609) These last are unfortunately not based on large numbers from all parts of the country, but their tendency to agree with each other and with the Army data gives one greater confidence in their trends Proctor's study (613, 614) is perhaps as good an illustration as any since it is longitudinal and covers one community He tested 1500 students in 1917-18 and ascertained their occupations thirteen years later When classified according to the occupational levels of their 1930 jobs, the results in Table 5 were obtained

TABLE 5

INTELLIGENCE IN HIGH SCHOOL AND OCCUPATIONAL LEVEL THIRTEEN YEARS LATER

Occupational Level	*Mean I Q.*
I Professional	115
II Managerial	108
III Clerical	104
IV Skilled	99
V Semi-Skilled	97

A final approach to the topic of occupational levels which should be mentioned is that in which the minimum intelligence required for success in the simplest type of employment has been investigated Most frequently referred to in this country is the study by Unger and Burr (887),

but Dunlop (222) made a similar study in Canada, and Abel (1), Beckham (58), Channing (153), Lord (481), Fairbanks (244), and others have also published on the subject, in the United States Table 6 lists typical occupations in which persons at the lower mental levels have successfully been employed after adequate induction on the job and when there were no serious personality problems to complicate things

TABLE 6

MINIMUM MENTAL AGES FOR SIMPLE OCCUPATIONS
(From Unger and Burr)

Mental Age	*Occupation*
5 years	Packing, garden work, scrubbing floors, simple washing
6 years	Light factory work, light domestic work
7 years	Assembly work, errands, pasting, farm work
8 years	Cutting, folding, garment machine operation, laundry, cooking
9 years	Hand sewing, press operation, filing, stock work
10 years	Routine clerical, general housework, machine operation, electrician's helper, painter
11 years	Selling, millinery work, janitorial work

One advantage in employing mentally handicapped adults in jobs such as the above is that, after the first period of careful supervision while they are learning the job, they are more likely to be satisfied with routine work and to be dependable employees than are other persons whose mental ability is such that they can legitimately aspire to more challenging work and are impelled to do so by boredom

Status within an occupation The multiplicity of occupations and the variety of criteria applicable to them have prevented any systematic study of the importance of intelligence for success *within* occupations, as contrasted with success *among* occupations or placement on the occupational scale But there have been a number of studies of the relationship between tested intelligence and success in certain specific occupations An examination of a few typical studies, of their results, and of the reasons for these results, is important to the user of intelligence tests in counseling and selection

Although the occupational level studies have shown that executives tend to make relatively high scores on intelligence tests, attempts to correlate intelligence and success in executive positions met with so little success during the 1920's that they fell into disrepute One such study was published in 1924 by Bingham and Davis (95) Using the army type intelligence test with 102 business executives, the correlation be-

tween test score and business success as indicated by a composite rating based on information contained in personal history records (salary, investments, debts, clubs, theatre attendance, etc), was — 10 Their conclusion that "superiority in intelligence, above a certain minimum (all were above the Army median), contributes relatively less to business success than does superiority in several non-intellectual traits of personality" has been generally accepted, and since the late '20's intelligence tests have generally been used only as a rough screening for executive positions However, Thompson (826, see also p 336) found a small group of superior executives superior to others on the Wonderlic Personnel Test

As in the case of executives, so in that of salesmen, the studies of the relationship between intelligence and sales ability have yielded negative results Most such studies have not been published, as they have been conducted by or for companies interested in their own personnel problems rather than by investigators with a more general interest But Moore (538 Ch 16) states that experience with salesmen of tangibles and of intangibles has led to an emphasis on work with tests of other types (largely interests, personality, and personal history) One typical study is reported by Anderson in his book on personnel work at Macy s (20) After administering the Otis Self-Administering Test of Mental Ability to 500 sales clerks, Anderson found that the distribution of intelligence scores clustered in the 80 to 110 range (75 percent), while 20 percent were below I Q 80 and 5 percent were above 110 This led to the conclusion that intelligence tests were of no value in selecting sales clerks, a conclusion reiterated by Anderson's successor (537 46) Actually, this approach seems too gross to be conclusive, a more refined analysis might, for instance, show that rug salesmen are and need to be more able than packaged food salesmen or girls who sell perfumes But this would be classification of sales jobs according to level, one would still need to ascertain whether the more intelligent rug salesman is more successful than the less intelligent rug salesman who also is above the critical minimum Such studies have not been published, partly for reasons given, partly because of the difficulty of obtaining enough comparable subjects in any one specialty for statistical study Perhaps they are not worth making, in view of what we know of the role of intelligence in occupations in which personality factors are of importance

Attempts to predict success in teaching, generally as evidenced in practice teaching while still a student, by means of intelligence tests have

met with the same lack of success as work with executives and salesmen Seagoe (688) made a study in which she correlated success in practice teaching, as rated on a specially constructed scale, with scores on a variety of tests, including the American Council Psychological Examination for College Freshmen She found no relationship between measured intelligence and rated teaching performance, although she did find some positive results in the area of personality Earlier studies found equally disappointing results with intelligence tests, but two recent co-ordinate investigations suggest that the situation may be more complex than this Rolfe (643) found no significant correlation between A C E scores and success in teaching in one- and two-room rural schools, whereas Rostker (652) reported a substantial relationship in larger schools with 7th and 8th grade pupils Apparently the occupation "teacher" is too broad a category for psychological study

Results in work with intelligence tests and clerical employees have been somewhat different, even though clerical workers are not, on the whole, as able intellectually as executives or teachers Some of the most convincing studies of this occupational group have been made by Bills of the Aetna Life Insurance Company, in collaboration, at times, with Pond of the Scovill Manufacturing Company (610) In the early study Bills tested 133 clerical employees at different levels of responsibility, and found a correlation coefficient of 22 with difficulty of the job Two and one-half years later the correlation was 41 for those who were still employed, the more intelligent having left the low grade jobs, often for advancement in the company, and the least able in the higher grade jobs having left them Aetna classified its office positions in 14 categories from A, low, to H, high Employees were classified also according to their intelligence test scores The results for a study of 903 employees in 1933 (610) showed that a clerical worker with a score above 100 had twice as good a chance of being promoted to a "responsible" position as an employee with a score of less than 80 At the same time they pointed out that almost as many employees with scores above 100 remain at the lowest levels as rise to the highest Another study has shown that intelligence is related not only to promotability in clerical work, but also to efficiency in the performance of clerical duties in a single job Hay gave a battery of tests to machine bookkeepers at the Pennsylvania Company, a Philadelphia bank (358) The operation was a routine, bimanual job, the criterion was production, that is, the number of debits and credits posted and of balances extended in a given amount of time, a criterion which had a reliability

of about 80 Hay points out that *amount* rather than accuracy is to be stressed, as inaccurate operators cannot keep their jobs The correlation between amount produced and Otis scores was 56 for 39 women operators This was higher than the coefficient for any other test in the battery, which included the Minnesota Vocational Test for Clerical Workers, Army Alpha, and several manual dexterity tests, although some of these had values independent of intelligence These results are reported to have been consistently obtained over a five-year period Unfortunately such studies are rare, and none are known to the writer which throw light on the applicability of the conclusion concerning intelligence and this one type of routine clerical work to other types of routine clerical work, although one would assume that success in semiautomatic tasks such as filing would, if a criterion were established, correlate even more highly with intelligence than does production in a practically automatic machine operation task

Pintner once wrote "The lower down the scale of industry we go, the less valuable do our present intelligence tests appear to be for the selection of workers" (604 489) He cited two studies, one by Otis (578) with a performance test administered to 400 workers many of them foreign born or illiterate, in a silk mill, and a study by Viteles (900) with motormen, in support of this statement Since that time a number of other studies have been made, in which more adequate statistical methods and better experimental design have been possible, with somewhat different results Blum and Candee (105) administered the Otis Self-Administering Test to 372 department-store packers and wrappers, while Forlano and Kirkpatrick (268) gave it to 20 radio tube mounters, the former finding that, although there was no relationship between test scores and production or supervisors' ratings for employees who had been on the job for some time, there was a suggestion of a relationship for new male employees, and the latter reporting that it was related to success only in the case of the less able learners the additional increment of intelligence was of no value to the superior beginners in learning a routine job Sartain (669) reported a correlation of 61 between refresher course ratings (reliability 77) and 46 aircraft factory inspectors' Otis intelligence test scores Shuman (716, 717) administered the Otis to inspectors, engine testers, machine operators, job setters, various types of supervisors, and other aircraft engine and propellor factory workers, the groups ranging in numbers from 25 to 99 each The correlations between Otis scores and supervisors' ratings (reliability 70 to 91) ranged from 39 to 57, depending upon the skill and

responsibility required by the job In view of results such as these, Pintner's conclusions from the earlier studies no longer seem correct Instead, the following conclusions concerning intelligence and success within an occupation seem warranted

1 People tend, in so far as circumstances permit, to gravitate toward jobs in which they have ability to compete successfully with others
2 Given intelligence above the minimum required for learning the occupation, be it executive work, teaching, packing, or light assembly work, additional increments of intelligence appear to have no special effect on an individual's success in that occupation This conclusion may be subject to revision as better criteria of success are developed, and *may not apply to more strictly intellectual jobs such as those in research or to some kinds of teaching*, but only to those in which personality and interest are peculiarly important
3 In routine occupations requiring speed and accuracy, whether clerical or semiskilled factory jobs, intelligence as measured by an alertness rather than a power test is related to success in the learning period and, in some vocations, after the initial adjustments are made

It should be noted that nothing has been reported on intelligence and success in the higher professions, in skilled trades, nor in unskilled occupations This is because no research on these problems has been located by the writer It seems likely that a positive relationship would be found in the first two, and none in the last, but this is still an unverified hypothesis

Job Satisfaction It has long been assumed that, even though a person might be able to do the work required by a job in which most of the workers are more able than he, the strain involved in keeping up with the competition would be such as to produce dissatisfaction in the worker It has similarly been widely held that ability considerably in excess of that required by a job causes dissatisfaction because of lack of challenge and consequent loss of interest in the work There is considerable clinical evidence to this effect, concerning both educational and vocational activities Pruette and Fryer (615) analyzed a number of case studies, confirming these beliefs for employed persons Scott, Clothier, and Mathewson (685 464) present charts showing the relationship between amount of school retardation upon leaving school (as a rough index of intelligence) and desire to change jobs in employees engaged in several different types of work in one company For 52 men employed in a repeti-

tive, monotonous inspection job the curve indicating percent desiring a change of job increased sharply with intelligence, in the simple but physically demanding foundry jobs the curve was bell shaped, the peak for the 42 men in question being at two or three years of retardation, with those more retarded or less retarded more likely to be satisfied, while in the assembly department, which offered a variety of somewhat more complex work, the curve for 86 men decreased with intelligence, for in this situation the abler men had more opportunity to use their ability and the less able felt the strain of difficult work Anderson (20 88–89) reported similar results in a study of labor turnover in the packing department at R H Macy's, where the brighter employees were found to leave their jobs sooner than the duller, seeking better outlets for their abilities

It is interesting to note that the studies referred to above were all made in the 1920's, when attention was focused on the use of the then new intelligence tests in personnel work Although such tests are still widely, and more discriminatingly, used for placement in business and industry, newer studies of the relationship between intelligence and job satisfaction do not appear in print This may probably be taken as an indication of the widespread acceptance of the relationship, but it is also due to the increased recognition of the fact that intelligence is only one among many complex factors in job satisfaction It would seem desirable, however, to supplement occupational norms of intelligence such as those compiled from Army data, which show the relationship between intelligence and usual occupation, with data on the relationship between intelligence and satisfaction in each occupation This would make possible the establishment of more adequate critical scores than would otherwise be possible Guidance and placement in terms of prospects of being able to compete with satisfaction as well as in terms of being able to hold a job has been shown to result in less instability, clinical evidence suggests that it also results in less irritability, aggression, self-recrimination, and escape into fantasy

Specific Tests

The Psychological Corporation's catalogue of tests recently listed 22 group tests of intelligence, most of them suitable for use at the adolescent and adult levels Even this partial list is obviously too long for adequate consideration in a volume such as this Annotated catalogues are available from publishers and distributors of tests, and brief critical reviews of

current tests appear periodically in the *Mental Measurements Yearbook* (126) There is a need, however, for a systematic review of the research which has been carried on with some of the widely used and more promising tests of intelligence, in order to provide the user with a clear picture of what has been done with these tests and with an understanding of their demonstrated values and limitations in vocational guidance and personnel work It is only upon such a foundation that tests can be used with maximum effectiveness and with minimum error In attempting to meet this need the writer's task is simplified by the fact that there are relatively few up-to-date tests of intelligence which have been widely used in vocational guidance and selection, the statistically analyzed results of which are to be found in the professional journals Even so, it seems wise to select a few representative tests and to treat them thoroughly rather than to cover all those which deserve to be included In this way space may be conserved and the repetition of similar findings for test after test avoided A few other tests are discussed more briefly and others are merely named Thorough coverage of a few representative instruments should provide the user of tests with insight into the nature and usefulness of the types of tests in question, and enable him to make his own evaluation of other tests in which he happens to be especially interested The selection of tests included in this or any other chapter, then, should be taken simply as an indication that they have been used in enough investigations for some facts concerning them to have accumulated, and as evidence of the author's preferences, rather than as a sign that these particular tests are necessarily intrinsically superior to certain others which are not treated in detail In deciding to use some other test, one should summarize all relevant data in a manner comparable to that of this book

The intelligence tests now used, whether individual or group, fall into three categories, which might be characterized as old type, new type, and factorial tests A brief discussion of these types should provide a useful orientation to the tests which are to be discussed in the following pages

Old type tests of intelligence consist of a variety of items arranged either in the spiral omnibus form or according to type with a time limit for each type, and yield only a total score or I Q The *Stanford-Binet* is an individual test of this type, the *Ohio State University Psychological Examination,* the *Henmon-Nelson Test of Mental Ability,* the *Pressey Classification* and *Verification Tests,* the *Terman-McNemar Test of Mental Ability,* the *Pintner General Ability Tests,* the various *Otis Tests,* the

Wonderlic Personnel Test, the *Army Alpha Test*, and the *Army General Classification Test* are group tests of the old type Although it is possible to analyze some of these tests in such a way as to obtain more refined estimates of the mental abilities of the persons tested, the tests were not designed for this purpose and they have no norms for the interpretation of such scores To point this out is not to deny the value of the overall score provided by any of these tests Of these, appreciable amounts of vocational validation data are available only for Army Alpha, the Army General Classification Test, the Pressey, and the Otis and Wonderlic Tests

New type tests include the same general type of items, but they are either arranged according to type in the test blank or rearranged in this way in the scoring process These grouped items provide a total score, as in the old type tests, but also part scores based on the type of item These part scores are generally *verbal* or linguistic and *performance* or quantitative The *Wechsler-Bellevue Intelligence Scale* is an individual test of this type, the *American Council on Education Psychological Examinations* and the *California Mental Maturity Tests* are group tests embodying the same features Norms are provided for linguistic and quantitative parts with the objective of making it possible to study the special mental abilities of the subject and to predict success in verbal or academic subjects, on the one hand, and quantitative or technical subjects on the other Differential occupational predictions were expected to be made possible by this type of special, as opposed to general, mental ability score A number of studies have been made of differential educational prediction on the basis of the A C E with conflicting results, these will be taken up in connection with this test Occupational evidence is still practically not available, the California and Wechsler tests still being relatively new and the A C E having been used largely in educational programs

Factorial tests of intelligence are still in an experimental stage, although the new type tests just described are based on the factor analysis work which preceded the development of factorial tests. The subtests which constitute a test of this type are included because they are heavily saturated with statistically isolated factors which seem to be fundamental components of intelligence Although, in combination, they measure what is commonly called general intelligence, factorial studies have shown that they are relatively independent of each other and unitary in nature Scores based on these subtests are therefore used as indices of special, or primary, mental abilities These are not as coarse as verbal or quantitative

ability, which factor analyses have shown to be constellations of abilities rather than unitary traits, but are more refined and include such verbal aptitudes as word fluency and verbal comprehension, and such quantitative aptitudes as spatial visualization and number facility The only published tests of this type are the Thurstone Tests of Primary Mental Abilities These will be discussed later as a promising technique still in the experimental stages, they cannot yet be said to have been proved useful

Two group tests of intelligence will be taken up in some detail in rounding out this chapter, and briefer discussions of three other tests will follow The two treated at length are the *Otis Self-Administering Test of Mental Ability* and its derivatives, and the *American Council on Education Psychological Examination for College Freshmen* The three to which less space is given are the *Army General Classification Test,* the *Thurstone Tests of Primary Mental Abilities,* and the *Wechsler-Bellevue Intelligence Scale*

The Otis Self-Administering Tests of Mental Ability (World Book Co , 1922)

The Otis Self-Administering Test was designed for use with senior high school and college students, and with adults Another form is suitable for elementary and junior high school students These have been revamped by Otis for special answer sheet and stencil scoring as the Otis Quick-Scoring Test of Mental Ability, and by Wonderlic as the Personnel Test, both essentially the same as the Otis S A with improved scoring techniques, improved time limits in the case of the Wonderlic, but less adequate norms in each case All three are widely used, the S A tests are described here as there are more data for them than for the other two tests

Applicability The Otis should not be relied upon with older college students and superior adults, as it is probably too easy As Otis' manual indicates, a number of investigations agree that when high school seniors and older persons are tested, it is preferable to use the twenty instead of the thirty minute time limit in order to correct for this weakness Older (573) has demonstrated, however, that the standing of persons tested with a twenty minute time limit should not be compared with that of persons tested with the longer limit

Contents There are 75 mixed items arranged in order of difficulty, some verbal, some arithmetical, and others spatial, they involve vocabulary, sentence meaning, proverbs, number series, analogies, etc A study

by Hovland and Wonderlic (384) reports that the arrangement of the items is no longer the best possible and that as many as 25 percent of the items are correctly answered by 90 percent of a large sample of adults (N=8300), for this reason the newer revisions are to be preferred as soon as adequate norms become available Crooks and Ferguson (183) found the items less suitable for college students than for adults, in both validity and difficulty level

Administration and Scoring There are no subtests to time, no special directions to give during the examination The time required is 20 or 30 minutes (see above) Scoring is by means of printed keys, and the score is the sum of the right answers

Norms The norms for the test are based on the distributions of scores for about 120,000 persons Raw scores may be converted into Binet mental ages derived from a combination of Herring-Binet scores and true mental ages as calculated from the distribution of raw scores by age groups This correction of Otis' data was deemed necessary because of the selective nature of the high-school groups used in standardizing the test

Bingham (94 338) has pointed out that Otis' college median is lower than that obtained by the College Entrance Examination Board As Otis' data come from a number of different colleges and represent all classes, whereas the Board's were obtained from a limited number of highly selective institutions, Otis' college norms are more nearly accurate That they do not err greatly on the easy side is shown by the fact that the average present day freshman makes an Otis I Q equivalent on the A C E Psychological Examination of 109 Otis' median college *student* I Q of 111 is equivalent to the 57th college *freshman* percentile on the A C E norms, probably a little lower than it would be if the lower-ranking freshmen had been eliminated Differences between colleges are great, so that local norms should be used in both counseling and selection Otis' manual reports median I Q 's for twenty-one colleges which range from 95 to 123

Factors Influencing Scores Baxter (52) administered the Otis to 48 college students and found that time and work-limit scores had an intercorrelation of 85, demonstrating that at that age level a *speed* score measures the quality of the work the subject can do Evidence has also been reported (99) indicating that college students who read poorly, as shown by the Iowa Silent Reading Test, are not underrated by the Otis Test, this was ascertained by comparing their Otis scores with their Army Beta (non-verbal) Test scores, a comparison which is vitiated by

the important common speed factor Scores have a very low negative correlation (−.03 to −.30) with age in adulthood (459)

Standardization and Initial Validation Otis' manual gives unusually complete and detailed information concerning standardization and initial, but little on subsequent, validation Many of the items in the tests were taken from existing instruments Preliminary editions were tried out on high-school groups of about 1000 each Items were retained if they distinguished clearly between superior young students and inferior older students in a given grade, the criterion of validity was therefore rapidity of school progress This suggests that the test, being academically standardized, might not be a very valid one for non-academic purposes Only occupational validation, and studies such as Hovland and Wonderlic's, can provide the answer The age and grade norms are based on large samples from various sections of the United States, not a random nor a stratified sample, but one large and varied enough so that to assume its adequacy seems sound the number for grade 6, for instance, is 15,715, for grade 12 it is 24,724, for college students, 2516 from 21 colleges These norms are those provided since publication of the test, utilizing additional data supplied by other investigators Strictly adult norms have not been published, despite widespread use at that age level

Reliability Forms A and B have an intercorrelation of .92 Reported reliability coefficients range from .90 to .97 (171) with the 20-minute, and of .86 with the 30-minute, limit with adults (577)

Validity Otis suggested in his manual that the method of standardization is the best indication of validity in an intelligence test This has already been described He also attempted validation through correlation with various criteria such as tests and grades

Correlations with grades in several high schools were .55, .57, and .59, the numbers ranging from 157 to 249 Segel (701) summarized six studies with nine coefficients ranging from .20 to .43 and a median of .38, while Hartson (348,349) found correlations with scholarship of .39, in high school and of .56 to .58 in college, Miller (532) found a correlation of .69 with high school grades, and one junior high school study (883) reported that the Otis test was the most useful of five tried The test clearly has a substantial relationship with educational achievement, one which varies as one might expect with the practices of the school, the marking systems of the teachers, and the range of ability and attitudes in pupils

Correlations with other tests are as follows the Otis had correlations of more than .70 with Army Alpha, the CAVD, and other tests (577),

Otis-Terman Group and Otis-Binet coefficients equal about 55 and 50 (532,577), the Otis I Q being 6 to 8 points lower than that obtained on the 1916 Stanford-Binet (150,855) Otis I Q's tend to differ from Binet I Q's, especially at the higher extreme These results are typical for this type of test, the Terman Group Test being anchored to the Binet is most like it, while the Otis, standardized without this base, is more closely correlated with group tests such as Army Alpha It is generally agreed that the use of the term I Q for converted Otis scores is not strictly justified Otis pointed this out in his manual, but used the term I Q because it is the standard method of measuring brightness, cautioning users of the test always to specify "Otis I Q" Despite the statistical impossibility of an adult I Q, the chronological age factor in the ratio of MA to CA having ceased to change after mid-adolescence, it is often convenient for test users to think in terms of I Q equivalents

Correlation with Success on the Job This topic has been dealt with at some length earlier in this chapter in connection with intelligence as measured by various tests A substantial number of the studies referred to in that section involved the use of the Otis Self-Administering Test, which together with Army Alpha in its various revisions was probably the most widely used test in business and industry during the 1920's and 1930's, especially at the clerical, skilled, and semiskilled levels In this section, therefore, only specific findings which may aid in the understanding and use of the Otis test will be mentioned

Hay and his associates have used the Otis in selecting bank clerks and calculating machine operators over a number of years at the Pennsylvania Company (359) There it has been found desirable to use 36 as a critical minimum raw score for clerical workers, with a 20-minute time limit, this is equal to a 30-minute raw score of 46, and an I Q of 104 When Otis scores were correlated with the production of machine bookkeepers (358) it yielded a coefficient of 56 (N equaled 39), figure which was sustained by subsequent experience

Shuman (716,717) has reported studies dealing with success in skilled employment He studied supervisors and skilled workers such as toolmaker learners and job setters, correlating Otis scores with ratings by supervisors The ratings had a reliability ranging from 70 to 91, the validity coefficients ranged from 39 to 57, increasing with the skill and degree of supervision exercised in the job Critical scores were established for each supervisory job, the minimum ranging from a raw score of 30 to one of 33 for foremen on the Otis Quick-Scoring, the I Q equivalent be-

ing 88 to 91, whereas that for inspectors in the same plants was 51 (I Q equals 109) Shuman calculated that the use of the Otis test would have improved the selection of excellent skilled and supervisory workers by from 15 to 20 percent Sartain (671) correlated Otis S A scores of groups of 40 foremen and 85 assistant foremen with supervisors' ratings, the reliability of which was as high as 79 or as low as 48 depending upon the comparison The validity coefficients were 04 and 16, other tests were no better

Studies of semiskilled jobs have been somewhat more numerous. Forlano and Kirkpatrick (268) analyzed the Otis test scores of 20 radio tube mounters, whose work requires considerable finger and hand dexterity Each worker was a new employee, tested upon application for work, each was rated "good" or "fair" by a supervisor after one month of employment There were as many fair as good employees among the group making above average or average scores on the Otis (I Q 95 or above), but six out of the seven employees who made below average scores (I Q 94 or less) on the Otis were considered fair and only one was considered good As the ratings were based on the induction and learning period, this suggests that, in semiskilled work, having more than the critical minimum of intelligence is desirable for rapid adjustment to the job, but that additional increments of ability are of little value It will be remembered that this was the sole positive finding of Blum and Candee (105,106) in their study of the role of intelligence in another semiskilled job, packing and wrapping here there was no relationship between Otis scores and production or supervisors' ratings for regular employees (those who had passed the learning period) and no relationship between intelligence and production in seasonal employees (whose brief employment period makes them learners for most of their period of employment), but the supervisors' ratings of the latter group did show a slight tendency for the superior male workers to be more intelligent than the inferior male workers The authors suggest that the failure to find a similar tendency among the women seasonal workers may be due to rating on a different basis In a project of the Office of Scientific Research and Development, Satter (673) found no relationship between Otis scores and submarine officers' ratings of enlisted men's performance

The Wonderlic Personnel Test, a revision of the Otis, was administered to 769 applicants for ordnance factory work, together with other tests, by McMurry and Johnson (500) The criterion of success in this study was supervisors' ratings of 587 employees still working when the follow-up

was made Although some of the tests did have rather high validity for some jobs, there were no significant correlations between intelligence and any of the ratings Tiffin and Greenly (846) administered the Otis to women electrical fixture and radio assembly workers, with similar results although other scores on tests were positively correlated with production, there was no relationship (23 ± 11) between intelligence and production As there was no analysis of the relationship during the learning period, it is impossible to draw any conclusions concerning the role of intelligence during induction into the job, but it is clear that, in these and in many other semiskilled jobs, intelligence is unrelated to success once the worker has made the initial adjustments

Success in skilled and semiskilled jobs has been correlated with Otis scores in training situations by Paterson and associates (588) and by Sartain (669) The former worked with junior high school boys, using a variety of criteria, several of which were occupational rather than educational in nature The Otis was administered, together with a variety of other tests, to 217 seventh and eighth grade boys, and correlated with instructors' ratings of the quality of work done in producing standard samples or projects in mechanical drawing and sheetmetal courses, and with an overall rating of the quality of their shop operations (N equalled 100 in this instance) These ratings were shown to have reliabilities of 87, 56, and 68, using the odd-even technique, and 93, 72, and 81 corrected by the Spearman-Brown formula The correlations with Otis scores were, respectively, 25, 16 to 19, and 21, although not high enough for use in counseling individuals, their relationships were statistically significant and indicate that intelligence plays some part in shop operations

Sartain's study (669), unlike Paterson's, used adult subjects in an industrial situation, but unfortunately his criterion was more educational than vocational in nature and a number of important details are not supplied He gave the Otis and other tests to 46 employees of the inspection department of an aircraft factory who were taking a refresher course for inspectors The sex and age of the employees are not described, although it is stated that many had considerable experience and some were relatively new in the department No information is provided as to the type of inspection work done failure of a given test to predict success in inspecting engine assemblies would, for example, mean something quite different from failure of a test to predict success in inspecting fuselages The two instructors rated each employee independently, their agreement being indicated by the unusually high correlation of 77, when the sub-

sequent merit ratings of 20 of these employees who were on the job a year later were averaged, the correlations beween instructors' ratings and merit ratings was 42 This suggests that the immediate criterion was not only fairly reliable, but also related to job success, even though based on performance in a refresher course rather than on the job. The correlation between Otis scores and instructors' ratings was 64, higher than that for any other test except the MacQuarrie Test for Mechanical Ability, other mechanical aptitude tests yielded coefficients of from 24 to 47 In another study of three groups of foremen (N=40, 53, 85) the criterion was supervisors' ratings (reliability= 79) but the validity of the Otis was only 04 to 16

Differentiation between Occupations Despite the widespread use with employed adults, no studies of intellectual differences between occupations have been made with the Otis test Shuman's study (717) established critical scores for certain jobs in one company, but these are of limited applicability Presumably occupational differentiation has been so well established with other tests, from which conclusions may be drawn for the Otis, that it has not seemed worth while to make such investigations It would certainly be impractical to try to improve upon the sampling of the Army testing in both World Wars, defective though it is in some respects

Job Satisfaction No studies have been located in which Otis scores have been related to satisfaction either in the current job or in the usual occupation The general paucity of work on this topic has already been discussed

Use of the Otis Tests in Counseling and Selection The evidence concerning the use of the Otis tests in educational counseling and selection clearly points to the conclusion that it is of value in estimating a given student's prospects of success in school or college Although many other factors need to be taken into account, and although the relationship between Otis scores and grades varies from school to school and from college to college, an individual's performance on such a test is one factor which should be known by that individual and by the counselor or admissions officer

Concerning its value in vocational guidance and selection, the evidence is not so clear But this is only to be expected, in view of the greater complexity of the occupational world and of the greater variety of demands made upon the worker by the various jobs in which he might engage Despite this fact, it has proved possible to establish critical minimum

Otis scores for employment in clerical, in skilled, and in semiskilled jobs, below which a disproportionately large number of workers fail and above which a reasonable proportion succeed, research with other tests indicates that this could also be done for executive and professional jobs

It has also been demonstrated that, at least in some semiskilled jobs, the Otis is valuable in predicting the speed and ease with which the new worker will make his initial adjustments to the job demands

Once the new worker has made the initial adjustment to a routine job, the Otis score has no value in predicting success either in terms of production or in terms of supervisory judgments At least one exception to this generalization is provided by machine bookkeeping, in which the work is routine but mental rather than manual, and demanding of great accuracy

No other generalizations concerning the Otis tests and vocational adjustment are warranted by the research However, certain other generalizations based on work with other intelligence tests which correlate reasonably well with the Otis are possible These have been discussed with the supporting evidence earlier in this chapter

Even if the results of studies of all intelligence tests and vocational adjustment are thus taken into account, there is a dearth of longitudinal studies of their predictive value in vocational guidance as contrasted with selection The vocational counselor must rely largely upon deduction and generalization from validation studies in selection programs and from cross-sectional studies such as those of the Army intelligence test data, and upon cautious insights which use a thorough understanding of the available research as a springboard for establishing working hypotheses More will be said on this subject in a later chapter on the interpretation of test results (Chapter 20)

The American Council on Education Psychological Examination (The American Council on Education, yearly)

Each fall the American Council on Education publishes a new form of its Psychological Examination for College Freshmen, an intelligence test used by some 300 colleges and universities L L and T G Thurstone of the University of Chicago have been responsible for the technical work on the tests, and the constant revision of forms which are used each year with thousands of entering college students has resulted in a superior series

Applicability Designed for and standardized on entering college

freshmen, the test may also be used with high school seniors, but the studies by Barnes (43) and Hunter (393) which have been made concerning changes in scores with increasing age have demonstrated a need for caution in making comparisons of high school students, older college students, or adults with the normative group In the latter investigation 87 of 105 college girls gained an average of 31 percentile points by their senior year, 75 percent of this change occurring during the first year The fact that published norms are in terms of college freshmen has tended to limit the use of the tests to that group, no tables are yet available to make possible the accurate interpretation of scores made by high school juniors or by college graduates

Contents Various editions of the test have included five or six sections such as sentence completion, artificial language, same-opposites (vocabulary), arithmetic reasoning, analogies (symbols, spatial), and number series, all grouped more recently into two parts to give a quantitative (arithmetic and spatial) and a linguistic as well as a total score The items are probably less affected by knowledge than those in most group tests, for the emphasis in selecting items was to choose those which measure ability to manipulate symbols rather than mastery of previously learned facts Thus in the artificial language test the subject is given a new vocabulary into which he must make translations, and in the analogies test he must pick out similarities and differences in unfamiliar symbols and forms As these tests and items have been selected and modified from earlier tests and tried out over a period of nearly twenty years on large numbers of subjects, with adequate funds for necessary research, they constitute an unusually valid and reliable instrument

Administration and Scoring Each subtest is preceded by a practice exercise, and both are closely timed The test requires about one hour all told Scoring is simple, machine-scoring methods being applied even in hand scoring

Norms Norms consist of percentiles for freshmen in liberal arts, teacher training, and junior colleges, a type of norm more helpful in the guidance of high school seniors planning further education than comparison with freshmen in colleges in general The numbers in each group tend to be about 60,000, 12,000, and 12,000 respectively It would be desirable in counseling concerning the choice of a college to have norms for specific institutions, in order to help choose one in which each student is most likely to succeed and to be satisfied Unfortunately, the need to "safeguard" the reputation of an institution keeps such data from be-

ing published, although in the long run each college would probably gain if it did declare its interest as being in students of a certain fairly broad mental level and in supplying a kind of education appropriate to that level College admissions officers use local norms for such tests as the A C E in evaluating doubtful candidates for admission The Thurstones have not supplied I Q equivalents because of the artificiality of adult mental ages, such equivalents are provided each year by the Educational Records Bureau and are helpful in interpreting A C E scores in terms useful in generalizing from college to vocational competition

In the absence of norms for specific institutions, the next best type would be norms for clearly defined and homogeneous groups of institutions The present classification of colleges into four-year, junior, teachers, and technical and professional colleges might seem at first glance to provide these, but as Crawford and Burnham (180 92–94) have pointed out this is not the case The four-year liberal arts colleges, for example, cover a range of scholastic aptitude which is almost as great as that of all four types of institutions (90 percent) They are, therefore, an extremely heterogeneous group, while the norms may be typical of colleges in general, the range is so great as not to be very helpful in counseling an individual about the choice of a specific institution Crawford and Burnham point out that the average Yale freshman is at the 90th percentile on the general norms, and nearly 80 percent of these freshmen exceed the national 75th percentile Norms should be provided for various classes of liberal arts colleges, adequately defined

Studies of sex differences reveal negligible differences in total scores, masculine superiority in quantitative parts, and feminine superiority in linguistic parts (840) This checks with data on interests reported by workers with Strong's interest inventory

Factors Influencing Scores Smith (723) has reported finding higher scores among urban than among rural students, as have other studies of urban-rural differences Whether this is primarily the long-term result of selective migration or the effect of environment and urban-constructed tests is still a question Barnes (42) found that two years of college mathematics had no appreciable effect on the Q scores of an experimental group of 40 students, when compared with 75 controls who had equal Q scores as entering freshmen but took no college work in mathematics

Standardization and Initial Validation New forms of the A C E tests are constructed so as to resemble earlier forms, although there are differences in details and innovations are gradually introduced as new

types of items are tried and adopted Each new form is thus based on extensive previous work which has proved its validity, in addition, it is administered for tentative standardization to 1000 or more students who have also taken the preceding form The scores of some 60,000 college freshmen who take the test each fall provide final norms Studies have occasionally been made to determine the academic predictive value of the examination and to establish its reliability The assumption is usually made, however, that since the new edition is anchored to the preceding editions and has similar norms it will be approximately as reliable and valid as they A report is published each year in the *American Council on Education Studies,* giving data on the form published the preceding fall

Reliability The reliability of the A C E tests has been consistently high One study by the test authors reported odd-even reliabilities of 95 for the total score, and of 87 and 95 for the Q and L scores respectively, for the 1938 college edition (840) Votaw (904) found a correlation of 74 between Otis scores in 7th grade and A C E scores six years later (N=70)

Validity It is generally accepted that one indication of the validity of an intelligence test is the carefulness of its standardization The care used in this series of tests is illustrated by subtest intercorrelations for the 1938 form which range from 30 to 65 with a median of 39 in an attempt to measure relatively distinct components of intelligence (840) The high reliability of part scores mentioned above is another illustration Another illustration is provided by the specific college norms reported by the authors (840) and by Traxler (858) who converted A C E scores to I Q equivalents and ascertained the median I Q 's of the freshmen in 323 colleges These ranged from 126 in a private liberal arts college to 87 in one junior college The median for liberal arts colleges was about 110, for teachers and junior colleges about 107 Schneidler and Berdie (680) have reviewed similar data As has been shown in numerous earlier studies, there is a college for almost every I Q level It is regrettable that they cannot be identified by professional counselors

Correlation with Other Tests The A C E test has frequently been correlated with other intelligence tests With the 1916 Binet a correlation of 69 (440) has been reported, while for the 1937 Revision it is 58, 62 (16) and 67 (507) With the Otis S A Higher Form, coefficients of 78 and 82 were found by Traxler (864) Hildreth found that the A C E gave approximately the same percentile ranks in the senior year of high school as the Binet had given previously to the same children in ele-

mentary school (372) Anderson and others (16) reported correlations of 48 and 53 between two different forms and the Wechsler-Bellevue, the two verbal scales are about as closely related (49, 51), but the performance and quantitative scales have less relationship (31, 39), a fact needing further investigation to make it clear just what types of concrete mental ability each of these scales measures Certainly it would be dangerous to interpret Wechsler-Bellevue performance I Q 's in terms of A C E Q-score validities, or vice-versa

The use of performance or quantitative scores in educational and vocational guidance is in any case still largely hypothetical, although in some selection programs specific evidence has been collected which makes possible the use of part scores The writer administered the 1938 college edition of the A C E to 123 high school juniors and seniors, together with the Nelson-Denny Reading Test, the Minnesota Vocational Test for Clerical Workers, and the Co-operative Survey Test in Mathematics (792) The results, shown in Table 7, indicate that A C E linguistic scores are more closely related to reading ability than are either quantitative or total scores, that linguistic scores predict achievement in mathematics as well as do quantitative scores, that linguistic scores are more closely related to name-checking scores than are quantitative but that they are equally related (or unrelated) to number-checking scores Traxler (863) found r's of 26 between Bennett Mechanical Comprehension scores and Q scores and of 34 between the same test and L scores Apparently the latter are a better measure of general ability than the former, and neither is a superior measure of special aptitudes It will be seen later that there is some evidence to support the belief that quantitative and linguistic scores have differential predictive value for college courses, but the evidence is conflicting, and data such as those just presented suggest that they are actually comparable in predictive value except for the closer relationship of linguistic scores and reading ability

TABLE 7

RELATIONSHIP OF A C E PART-SCORES TO OTHER ABILITIES

A C E	*Reading*	*Mathematics*	*Name-Checking*	*Number-Checking*	*T*	*Q*	*L*
Total	66	65	62	26	—	75	92
Q	37	56	41	18	75	—	—
L	80	56	58	22	92	47	—

Bryan (123) and Estes (240) have reported correlations of 05 to 36 and 45 between Q scores and the Minnesota Paper Form Board (Revised)

In the former study, the spatial subtests correlated .55 with the Paper Form Board. This is a lower correlation than is generally found between intelligence tests and the Paper Form Board (p. 301), perhaps because of the homogeneous population.

A totally different line of investigation was opened up by Munroe in a study of the relationship between A C E scores and Rorschach indices (553). She administered both tests to 80 students at Sarah Lawrence College, and ascertained the difference between the Q and L percentiles for each girl. These difference scores were distributed, and the top and bottom quartiles were selected for further study. This gave Munroe one group of "higher L's" and one of "higher Q's." The Rorschach patterns of each of these groups were then analyzed and contrasted, with the following conclusions:

There were no differences in general adjustment as measured by the Rorschach Inspection Technique.

There were no differences in the number of responses nor in the number of words in the protocols of the two groups.

The *higher Q's* gave a significantly larger percentage of responses in which form was the determinant.

The *higher Q's* gave significantly more accurate form responses.

The *higher L's* gave significantly more movement responses.

The personality picture obtained from the above data is one of a subjective, imaginative, *higher L* syndrome and of a more objective, literal, outer-reality-bound *higher Q* syndrome. The latter type (if persons at the extreme of a continuum may be called that) resembles that found in paleontologists by Roe (636) and described in a later chapter. In pointing this fact out, Roe also states that the *higher Q's* were found to choose more scientific courses than the *higher L's*. If these findings are confirmed by other studies it would seem that differences in quantitative and linguistic scores may be indicative of differences in the utilization of intelligence arising from differences in personality, as well as, or perhaps even rather than, differences in primary mental abilities. Such a radical conclusion would be compatible with the findings that Q and L scores are not differentially related to success in quantitative and linguistic subjects, and are related to the choice of one or the other type of curriculum. It would not fit in with contemporary factor theory.

Correlation with Grades The relationship with achievement has been most intensively studied, academic prediction being the purpose of the test Studies from various earlier editions yielded validity coefficients ranging from 17 to 81 for grade-point averages (284) and from 34 to 60 with freshmen marks, and correlations of 43, 43, and 54 (284,456, 494,495,632,705) with long-term averages of groups of 228, 1,052 and 378 students Subsequent studies reported correlations of 39 to 60 with grades in various colleges (16,427,495,553), the mode being about 55 Modal correlations with first semester grades are about 45 for engineers and 50 for art students For grades over four years the correlations are about 45

Weintraub and Salley (915) found that, at Hunter College, 14 percent of the upper half of a freshman class of 1064 students were dropped for poor scholarship over the four-year course, as contrasted with 24 percent in the lower half on the A C E The range of intelligence in this group was of course limited

At the University of Chicago (840) correlations with introductory biology marks ranged from 43 to 47, humanities, 46 to 53, physical science, 39 to 46, social science, 46 to 51 (N=200 to 2000) Slightly (2 to 6 pts) higher results were reported by Shanner and Kuder (712) The correlation with marks for students of agriculture in another institution was 49, engineering, 45, general, 49 (546) This appears contrary to the suggestion of some that the test should be more valid in liberal arts than in other colleges

Part-scores have been related to achievement in specific subjects and fields by several investigators Segel and Gerberich (4) correlated part-scores with marks in English, foreign languages, and mathematics, with the results shown in the left-hand column of each pair in Table 8 Coefficients for variables which should theoretically be highly correlated are shown in italics

In another study by the same authors (704) part scores were correlated

TABLE 8

CORRELATION BETWEEN A C L PART-SCORES AND ACHIEVEMENT

	English		*Foreign Languages*		*Mathematics*	
A C E	*Marks*	*Test*	*Marks*	*Test*	*Marks*	*Test*
Completion	*41*	*44*	*19*	*55*	33	31
Art Lang	*54*	*65*	*38*	*75*	43	38
Analogies	38	38	25	44	*31*	*41*
Arithmetic	20	28	14	38	*38*	*62*
Same-Opposites	*51*	*52*	*32*	*70*	36	35

with Iowa Placement Test scores in the same subjects The results appear in the right hand member of each pair of columns in Table 8 The discrepancies are such as to be surprising were it not for the unreliability of marks, nevertheless, patterns of ability and achievement seem to exist, the verbal tests being more closely related to the verbal subjects, the quantitative tests (in one case) to the quantitative subjects Similar curricular relationships were later found at the University of Florida (546) Work such as this, combined with Thurstone's factor analysis (840) led to the use of Q and L scores in more recent editions Evidence which indicates a need for caution was published by the writer (792), to the effect that whereas the total scores on the A C E test correlate 65 with the Co-operative Survey Test of Mathematics, both the Q and L scores correlate 56 with the same test, Q scores giving a prediction of achievement in mathematics in no way superior to that yielded by L scores On the other hand, while the total score has a correlation of 66 with the Nelson-Denny Reading Test, that for Q is 37 and that for L is 80, indicating a genuine difference in Q and L scores Generally similar results have been obtained by four other investigators using grades as criteria (16,42,503,764) MacPhail's study (503) involved analyses of data at both secondary and collegiate levels, the latter were treated in terms of both curriculum and courses Representative data from two of his tables are reproduced in Table 9

TABLE 9

CORRELATION OF GRADES IN QUANTITATIVE AND LINGUISTIC COURSES WITH Q AND L SCORES ON THE A C E, AFTER MACPHAIL

N	*Courses (Quantit.)*	*Q*	*L*	*C R*	*N*	*Courses (Linguistic)*	*Q*	*L*	*C R*
49	Descriptive Geom	39	13	2 10	95	French (Intermed.)	— 02	45	5 26
64	Chemistry (Elem.)	665	50	2 06	53	German (Elem.)	13	20	0 54
31	Chemistry (Qual An.)	19	01	1 09	27	History (U S)	14	50	2 24
82	Mathematics (Trig., Cal.)	31	21	0 98	48	History (Europe)	38	44	0 49

Of the courses for which data are not reproduced here, only psychology among the "quantitative" subjects showed a possibly significant difference between the correlations, and that was in favor of the L score, there were no significant differences among the other "linguistic" courses The conclusion drawn by MacPhail is that data of this type must be obtained by each institution if it wishes to use Q and L scores for selection and guidance, certainly any blanket use of such scores in counseling is now unwarranted, and, if one were to generalize from his study (as adequate

as any now available), it would be to the effect that L scores are as satisfactory as Q scores for predicting success in mathematical and scientific courses, and perhaps slightly more satisfactory for predicting achievement in some linguistic and verbal courses

Estes (240) correlated A C E scores with grades in analytic geometry for 76 engineering freshmen with the following results r Q and grades= 33, r L and grades= 15 This agrees with MacPhail's findings Bryan (123) found correlations between A C E scores and art grades varying from 02 to 37 for various types of art students (N=1008), those for the quantitative parts tending to be slightly lower than those for the verbal, but the trends are not significant

Part scores on tests such as this presumably measure constellations of primary abilities, as Thurstone (840) has shown, although Munroe s exploratory work on personality relationships (p 119) raises important questions These *may* be related to achievement in special fields as reported by some investigators, but it is obvious that more conclusive evidence is needed before A C E Q and L scores are relied on in differential prediction or counseling

Correlation with Success on the Job It is to be regretted, in view of its excellent construction, widespread use and the extensive information on hand concerning it, that the A C E test has not been adequately validated for vocational guidance and selection at the business and professional levels There are practically no validation studies of this test using strictly vocational criteria, although several studies have shown that its total scores are related to success in some types of professional training, e g, engineering (191), and, in some institutions, nursing (619) Seagoe (688) found that well-adjusted student-teachers, and maladjusted student-teachers of average or low intelligence in one college, tended to remain in training whereas the bright but maladjusted students dropped out—perhaps because they recognized the misfit and saw other more appropriate opportunities Ratings of success in practice teaching did not correlate significantly with A C E scores Rolfe (613) found no relationship (r = — 10) between A C E scores and the teaching success of 52 Wisconsin one- and two-room school teachers, the criterion being tested pupil progress Rostker (652), however, applying similar techniques to 28 teachers of 375 seventh and eighth grade pupils found a correlation of 57 Perhaps teaching in larger schools is a more intellectual activity Bransford and others (117) found a correlation of 64 between A C E scores and ratings of the administrative effectiveness of 20 civil servants at the top

management level These findings suggest that intelligence as measured by the A C E plays a part in the intellectual aspects of some vocations, including those important in training, but that in other occupations, whether in training or in practice, other factors are more important

Differentiation between Occupations Two studies have found the usual relationship between parental occupation and student intelligence. Byrns and Henmon (129) found significant differences between adjacent occupational levels, except the business and clerical and the skilled and semiskilled Smith's study (722), based on 5487 students, found similar differences

Job Satisfaction No studies with this test have been located which bear, directly or indirectly, on job satisfaction although Berdie (78) showed that A C E scores were not related to satisfaction with training in engineering

Use of the A C E Psychological Examination in Counseling and Selection This review of the A C E Psychological Examination shows that it has been studied in most of the ways in which other tests have been tried, although rarely in investigations of vocational adjustment There is probably more material concerning its educational significance than there is for any other single test It is a reliable and valid test of scholastic aptitude or general intelligence at the college level The test goes beyond this, however, in attempting to break down the concept of "general intelligence" by providing part-scores for what logical and statistical analysis indicate may be special aspects of intelligence As Thurstone (840) has shown in a factor analysis of the 1938 edition, these aspects of intelligence are not primary abilities or factors, but constellations of related factors This breakdown is thus a compromise attempt to take advantage of the findings of factor analysis and yet to provide a practical measure for administrative and guidance use It is promising because it represents a step in advance in group testing technique without departing so far from proved techniques as to make it a purely research instrument, but its part-scores are still of uncertain value in differential diagnosis and prediction

The freshman college norms are perhaps the most adequate available It is unfortunate that the same forms are not standardized at other educational and age levels, and that its vocational significance is not better established However, the high correlation with other intelligence tests, together with the equivalent scores which have been made available, make it possible cautiously to use occupational and educational norms

established for other tests It should be remembered in so doing that Otis I Q 's are not the same as Binet I Q 's because of different methods of calculation, that I Q 's are artificial equivalents and not true ratios of mental to chronological age at older adolescent and adult levels, and that equivalent scores are based on averages and may therefore be distorted in extreme cases

The Army General Classification Test (The Adjutant General's Office, War Department, 1940, Science Research Associates, 1947)

This test was devised by the Adjutant General's Office when Selective Service was adopted in 1940, as a substitute for the widely used Army Alpha of World War I The two orginal forms designated by the Army as AGCT-1a and AGCT-1b were used from October 1940 and April 1941, respectively, to October 1941 The two final forms, AGCT-1c and AGCT-1d, were equated with the first two and were used in the testing of all men and women who were inducted into the Army between October 1941 and April 1945 AGCT-1 was administered to a total of well over 9,000,000 persons It was so widely used that more than 4000 persons daily were tested With the introduction of a completely revised classification test, based on more modern principles of intelligence test construction and yielding separate scores for verbal, numerical, and spatial aptitudes, forms 1c and 1d became obsolete The large number of men and women who had been tested with these forms, and the vast amount of educational and occupational validation data which had been accumulated for them, made them unique in the history of psychological or vocational testing Two forms were therefore released for civilian use, the first civilian edition appearing as forms AH (hand scored) and AM (machine scored) This, it should be noted, is AGCT-1a, which is not the widely used Army form, but its predecessor, to which the widely used forms, 1c and 1d, were calibrated

Applicability The AGCT was designed for use with draftees, that is, with young men between the ages of 18 or 20 and 36, with widely varying amounts and types of education, and with even greater differences in general cultural background In order to make the test applicable to this group, an attempt was made to avoid items which might be greatly influenced by schooling beyond the first few grades and by other cultural inequalities Information items were not used Instead, vocabulary, everyday arithmetic, and spatial items were included A special effort was made to make the items seem sensible to young men from all walks of

life The data on distributions of test scores, for example the occupational norms to be discussed later, indicate that the objective of getting a wide-range intelligence test of reasonable brevity was achieved It was used also with young women who volunteered for the Army, and the data on such groups give no reasons for questioning its applicability to women Observation of the use of the test with both men and women suggests that they find the types of items acceptable, although the block-counting sections apparently make a special impression the test is often referred to by examinees as "that test with block counting in it"

As military experience showed that young men and women with widely varying amounts of education seemed to be able to manage this test, it seems likely that it could be used also in the last years of high school However, no objective evidence on this point has as yet been published Its use with older people might be questioned, for although the correlation with age in a representative enlisted population was only 02, it was −33 and −20 for two groups of officers which included many men who were older than most draftees (736,737) As pointed out in the official report, this is probably due to the influence of the speed factor, although an attempt had been made to minimize that by a time limit in which all examinees could, if not finish, at least show their power

Content The test consists of three parts vocabulary, arithmetic problems, and block counting Three practice parts introduce the test to insure familiarity with the procedure A sample vocabulary item is "To permit is to, a) demand, b) thank, c) allow, d) charge" The arithmetic problems involve real life situations, such as dividing rounds of ammunition among a group of men, finding out how many more cows one man has in comparison with his neighbor, and computing the amount of money each man on a baseball team would have to contribute in order to supplement the club's treasury in buying uniforms The block-counting items are of the familiar type, like those used in the MacQuarrie

There are 30 practice items in the civilian edition, and 150 test items, in contrast with 10 practice items and 140 test items in Army editions 1c and 1d The manual does not indicate which Army form was used, but this fact suggests that it is one of the two older forms (actually Form 1a, confirmed in a letter from John Yale of Science Research Associates, dated April 14, 1948) AGCT-1a was standardized on 2675 men aged 20–29 Form 1b was standardized on 3856 men who also took Form 1a, in 1941 The correlation between scores on the two tests was found to be 95, and their means and standard deviations were practically identi-

cal Forms 1c and 1d were prepared immediately after 1b, administered to 1782 men, and compared with 1a The two new forms were found to be somewhat more difficult than 1a, and somewhat more discriminating in the upper ranges (736 763), no comparisons were made with form 1b, but presumably the same would be true of it

Administration and Scoring The testing time is 40 minutes Directions in each booklet are complete, making the test self-administering The civilian edition uses the step-back format, in which each page is slightly narrower than the one before it and the answers are recorded on successively exposed columns of the answer sheet This has the great advantage of making a manageable booklet and answer sheet, and of minimizing recording errors The hand-scored form provides the examinee with a pin with which to prick holes in the answer sheet instead of marking it The holes which appear in marked areas of the back of the answer sheet are counted and indicate the number of right answers Scoring takes only about one minute per test Raw scores are converted into standard scores known as Army standard scores, for which the mean was intended to be 100 and the standard deviation 20 These can also be converted into percentiles, a table in the manual being provided for this purpose

Norms As the extensive Army norms are for AGCT-1c and 1d (more than 8,000,000 men), it is to be regretted that the civilian form is one of the preliminary editions As they are very similar, even though not identical, it may be safe to use the general norms

The manual provides a table for the conversion of raw scores into Army standard scores and percentiles There is no indication as to what size or type of group this table is based on It is military, but whether or not it is the standardization group for the same form, or a much larger group tested with equated forms is uncertain A sentence elsewhere in the manual indicates that it is based on 160,000 (undescribed) inductees The mean raw score of the standardization group used with form 1a was 78, which gives a standard score of 102 and a percentile of 45 according to the manual, the percentile would be the 50th if the standardization group were the norm group As the manual's norms are for a larger number of persons than were tested with this form, data from other forms which had been calibrated with this one must have been used Such matters should be made clear in the manual or in accompanying publications

Occupational norms, in the form of bars representing the middle 50 and 80 percents of each of 120-odd occupations, are also included in the

manual Again, it is not clear what forms of the test were administered to these groups If 1c and 1d were used, the norms may not be strictly applicable to the civilian form (AGCT-1a), which was found to be easier and less discriminating at the upper levels Persons of average or high-average ability would seem more able to compete in executive and professional work than they actually are The means in the manual's occupational norms are almost identical with those of the longer list of occupations covered by Stewart's analysis (758), but the numbers of cases are in some instances smaller, and in some larger, than hers

Standardization The standardization of the various forms of the AGCT has been described in readily available journals by the staff which developed it (736,737) and need not be repeated here Steps which should be noted include the fact that a large item-pool was developed, and the seemingly most appropriate items were selected from it, each successive form was equated with the previous forms (but as noted previously 1a and 1b were somewhat easier than 1c and 1d) the estimated mean of the first form proved to be too low, so that when the calibrated scores of later forms were standardized the actual mean standard score was between 100 and 110 rather than 100, one sample of more than 91,000 men having a mean of 105 (758 31)

The *reliability* of the various forms was ascertained, the retest reliability with varying intervals between tests being 82, the alternate-form reliability between 89 and 95, the Kuder-Richardson reliability between 94 and 97, and the corrected odd-even reliability 97 (736 765) These are quite satisfactory

Validity As the AGCT was devised as a measure of learning ability and routinely administered to all enlisted men and women in the Army, it was used as a predictor of success in training for many types of specialties But it was also possible to relate scores on this test to certain criteria from the previous civilian experience of the persons tested, such as the amount of education they had obtained (it having been well established by other studies that brighter people tend to get more education) and civilian occupation (it has been seen that occupations can be ranked according to their intellectual requirements)

Education, as measured by the highest grade attained, was correlated with the AGCT scores of 4330 men, the coefficient being 73 (736) This may be unduly high, because socio-economic status is correlated with each of these variables, but it is an indication that the test has some of the validity which has generally characterized good intelligence tests

Tests of intelligence which have been correlated with the AGCT include Army Alpha, Otis S A , and the American Council on Education Psychological Examination (736) The most representative populations for which such data have been published ranged in numbers from 750 to 1646 The correlations were 90 for Army Alpha, 83 for the Otis, and 79 for the A C E

Other tests with which the AGCT has been correlated include those used in the selection of aviation cadets (214 Table 5 9) The correlation with a test of *reading comprehension* was 53, *mechanical comprehension* 32, and *mathematics* 45 The correlations with tests of *manual dexterity, co-ordination,* and similar capacities were generally below 20 These data were obtained from a group of more than 1000 unselected applicants for cadet training

Success in training was the most commonly used criterion for the validation of the AGCT A summary of such results was compiled by the staff of the Personnel Research Section of the Adjutant General's Office (736), and is reproduced here with additional data from DuBois (214) The means and sigmas of the various military training groups are given, together with the correlations with the criteria As the authors point out, preselection of students, sometimes on the basis of this test, makes the relationship seem lower than it actually is, in some instances, whereas in others the true relationship is shown Motor mechanics, for example, were not preselected, and r equalled 69, teletype maintenance students were preselected, and in their case r equalled 20 It would be necessary to sort these data into at least two groups, according to whether or not they had been preselected, in order to generalize concerning the types of training in which the test best predicted success Even then, it would be necessary to be cautious, because of the presumably academic nature of much of the training, even for specialities which were very concrete and practical The example of Navy aerial gunnery training has been cited elsewhere (p 35) as evidence of the fact that intelligence tests sometimes predict success in training because the training is unnecessarily abstract, and that when the training is made more life-like intelligence tests lose their predictive value

It is worthy of note, as the AGO authors pointed out, that the correlations between AGCT and grades in Army Specialized Training (college courses) and also in most West Point courses, tend to be low They range from 12 to 40 The authors point out that this is no doubt partly due to the extreme preselection which had taken place in both pro-

TABLE 10

VALIDITY COEFFICIENTS OF THE AGCT

Population	*Criterion*	*N*	*Mean*	*SD*	*r*
Administrative Clerical Trainees, AAF	Grades	2947	121 7	11 1	40
Clerical Trainees, AAF	Grades (weighted)	123	125 9	9 9	44
Clerical Trainees Armored	Grades	119	125 3	8 3	33
Clerical Trainees, WAAC	Grades	199	116 8	12 0	62
Airplane Mechanic Trainees	Grades	99	104 8	10 6	38
Airplane Mechanic Trainees	Grades	3081	118 1	10 7	35
Motor Mechanic Trainees	Grades	318	88 3	24 4	69
Tank Mechanic Trainees	Grades	237	116 6	11 3	39
Aircraft Armorer Trainees	Grades	1907	117 3	10 9	40
Aircraft Armorer Trainees	Ratings	439	112 7	12 1	27
Aircraft Welding Trainees	Grades	583	114 8	10 3	26
Bombsight Maintenance Trainees	Grades	105	129 1	10 5	31
Sheetmetal Trainees AAF	Grades	764	115 6	10 3	27
Teletype Maintenance Trainees AAF	Grades	487	121 5	12 1	20
Radio Operator & Mechanic Trainees AAF	Grades	1055	122 4	11 1	32
Radio Operator & Mechanic Trainees AAF	Code Rec Speed WPM	217	117 4	11 7	24
Radio Operator Trainees WAAC	Grades	154	116 2	11 7	38
Radio Mechanic Trainees AAF	Grades	419	108 2	13 0	49
Gunnery Trainees Armored	Grades	66	120 0	12 1	50
Field Artillery Trainees, Instrument and Survey	Grades	68	104 7	6 5	33
Motor Transport Trainees, WAAC	Grades	269	111 4	13 6	31
Tank Driver Trainees	Ratings	130	87 7	19 5	16
Truck Driver Trainees	Road Test Ratings	421	95 5	20 1	13
Bombardier Trainees AAF	Grades Academic	40	111 5	18 6	62
Aircraft Warning Trainees Plotter Teller	Grades Theory	119	107 1	15 6	73
Aircraft Warning Trainees, Plotter Teller	Grades Performance	119	107 1	15 6	26
Intelligence Trainees, AAF	Grades, Academic	1 4	118 9	10 6	51
Photography Trainees AAF	Grades	431	123 0	11 9	24
Cryptography Trainees AAF	Grades, Phase 1	117	129 9	9 7	31
Weather Observer Trainees AAF	Grades	1042	130 2	12 5	43
Aviation Cadets Experimental Group	Pilot training Pass fail	1080	113 0	13 8	31*
Officer Candidates Infantry	Grades Academic	103	123 0	10 8	30
Officer Candidates, Ordnance	Grades, Academic	190	128 2	9 6	41
Officer Candidates, Signal Corps	Grades Academic	213	128 6	10 1	36
Officer Candidates, Tank Destroyers	Grades Academic	52	125 8	10 7	44
Officer Candidates, Transportation Corps	Grades Academic	314	126 4	9 8	38
Officer Candidates, WAAC	Grades, Academic	787	128 4	11 3	46
Officer Candidates Infantry	Leadership Ratings	201	122 6	10 8	12
Officer Candidates, Ordnance	Leadership Ratings	190	128 2	9 6	09
Officer Candidates, 13 Arms & Services	Success vs Failure	5186	128 7	10 0	28*
AST Trainees, basic engineering	Grades, Inorganic Chemistry	222	126 6	7 8	21
AST Trainees basic engineering	Grades Math (Trig)	222	126 6	7 8	16
AST Trainees, personnel psychology	Ranks in Statistics	132	134 2	10 4	25
AST Trainees, personnel psychology	Ranks in Tests & Measurements	130	134 0	10 3	29
West Point Cadets, 4th Class	Grades, English**	932	131 3	10 9	40
West Point Cadets 4th Class	Grades Mathematics**	932	131 3	10 9	43
West Point Cadets, 4th Class	Grades Military Topography	932	131 3	10 9	40
West Point Cadets 4th Class	Grades Tactics	932	131 3	10 9	29
West Point Cadets, 4th Class	Grades French**	167	130 2	11 0	22
West Point Cadets, 4th Class	Grades, German**	164	132 4	10 9	20
West Point Cadets 4th Class	Grades, Spanish**	92	131 3	10 9	19
West Point Cadets, 4th Class	Grades Portuguese**	168	130 0	10 3	12

* Biserial Correlation
** First Term
After Staff, AGO (736) by permission of the American Psychological Association and DuBois (214)

grams Despite this, however, the correlations with grades in English and Mathematics at West Point were 40 and 43 Strong (776) has pointed out another reason for the poorer predictions in specialized training, namely, the fact that a substantial number of men were sent to training in which they had little genuine interest, either because they thought it

would be a pleasant type of assignment or because quotas had to be filled With motivation undermined in this latter way, the correlation between ability and grades would be definitely lowered

The value of the AGCT as a predictor of success in pilot training can be ascertained by comparing it with the tests of the Aviation Cadet Selection program It is obviously not relevant to compare it with tests of special aptitude, interest, or temperament, but it may legitimately be compared with the general qualifying examination administered to applicants for preliminary screening, in order to ascertain the relative value of general intelligence tests and of custom built tests of ability to adapt to the learning requirements of a specific training program Table 10 has shown that in the experimental group of more than 1000 cadets sent to pilot training regardless of test scores the AGCT had a validity of 31 with a pass-fail criterion For this same group, with the same criterion, a test of learning ability designed with flying training specifically in mind had a validity of 50 (214 191) The pilot stanine (weighted combination of special aptitude test scores) had a validity of 66 Obviously, although the general intelligence test had some value for predicting success in pilot training, it did not measure certain factors which were of considerable importance and which were tapped by the more specialized tests

Occupational differences have been studied with the AGCT as with Army Alpha, but so far only for a one percent sample of the tested population Some of the data for this test are presented in Table 4, on pages 96–97 Stewart's paper (758) has shown that, as in the case of World War I data, occupations can be ranked according to a hierarchy of intelligence, there is considerable overlapping of occupational groups, and the spread of intelligence is greater in the lower level (less selective) than in the higher-level (more selective) occupations It is worth noting that although 90 percent of the highest ranking occupational group in either sample, accountants, made scores of 114 or better, more than 10 percent of the men in the least able occupational group, lumberjacks, made equally high scores The overlapping is even greater among the occupations which are nearer to the middle of the distribution Scores on this, as on other, intelligence tests can therefore give only a very general indication of the occupational level at which a person might best aim, notwithstanding the great variety of available occupational norms which seem to indicate the contrary

Stewart's analysis compares occupational ranks in World War II

with those found in World War I data She found that only gunsmith, toolmaker, machinist, telephone and telegraph lineman, locomotive fireman, meat cutter, and boilermaker had made appreciable gains in position relative to other occupations Occupations which had lost status were draftsman, file clerk, electrician, auto mechanic, pipe fitter, auto serviceman, chauffeur, and motorcyclist As Stewart points out, it is difficult to know just how to interpret these differences, or the relative lack of differences, between the two sets of norms The sampling of occupations during the two wars may have been different certainly selective service did not operate on the same principles, and some occupations may have been granted deferments more liberally in one war than in the other because of differing industrial needs This would result in inferior members of an occupation being its representatives in the war in which their group was considered essential to the civilian war effort In the absence of detailed information on the basis of which corrections in the occupational means and deviations can be made, one can use the Army occupational intelligence data only as a very rough guide

A seemingly sound form in which AGCT occupational norms have been presented for this type of use is the table prepared by Stewart and reproduced earlier in this chapter, in the discussion of occupational intelligence levels (pp 96–97) In this table will be found broad groupings of occupations on the basis of the AGCT scores characterizing their members This arrangement minimizes the likelihood that undue emphasis will be placed upon insignificant differences within a level, but at the same time it risks overemphasizing the importance of differences between top and bottom occupations in adjacent levels One wonders, for example, whether the differences between chemists and lawyers are as great as the fact that one falls in Stewart's highest group and the other in her next highest group implies The difference is, actually, one of three AGCT score points, or less than one-fifth sigma Although the writer has used such tables, and reproduced one based on Army Alpha in an earlier text (793 56), it now seems wiser to work from a graph such as that provided in the manual The scaled arrangement permits the counselor and client to study broad groupings by drawing lines wherever they may wish, and at the same time encourages the realistic consideration of overlapping and of relative standing in a variety of occupations Data for a longer list of occupations will be found in the Stewart reference (758 Table I)

Use of the AGCT in Counseling and Selection It is clear from the

relationship between the Army General Classification Test and other standard tests of intelligence that this instrument is a measure of learning ability This conclusion is reinforced by the consistently significant correlations between AGCT scores and success in administrative, clerical, mechanical, electrical, academic and other more specialized types of training in the Army, even though the nature of the data did not permit generalization concerning its relative importance in each of these types of training

Although no evidence is available concerning the relationship between AGCT scores and occupational success, the data on differences between occupational groups have been seen to confirm the opinion that persons with higher AGCT scores are likely to make good in higher level occupations The details are in general agreement with the findings of studies made with other tests, so that generalizations can probably be made from this test as from other standard tests of intelligence These would be to the effect that those with high scores are most likely to master new jobs rapidly, to rise to positions of responsibility, and to be satisfied in high level occupations

The test can be used in high schools, colleges, guidance centers serving adolescents and adults, employment offices, and business and industrial establishments It is perhaps unfortunate that the name "Army" has been kept on the test booklets (although it should be identified correctly among professional users), as this may injure rapport with some subjects Experience will no doubt throw more light on this problem The contents and form are quite appropriate despite some items dealing with military objects or situations The occupational norms make the test useful for vocational counseling, and for selection in the absence of local norms The lack of college student norms makes it less useful than the A C E , Otis, and certain other tests for educational guidance, but this defect is to some extent remedied by the availability of means for certain special types of college students, and by the substantial correlations found with grades in various types of training courses

The Thurstone Tests of Primary Mental Abilities (American Council on Education, 1938, 1941, Science Research Associates, 1947)

The Tests of Primary Mental Abilities were developed by the Thurstones in an attempt to provide practical batteries of tests implementing their work in the isolation of primary mental abilities The "Chicago" (long-form, two hours) and "SRA" (short-form, 45 minutes) Tests were

designed for use primarily at the high school level (845), another battery has been added for the lower age levels Only the long experimental and "Chicago" forms are discussed here, as there are very few data concerning the short forms

Description

The *Chicago* tests were standardized on children in the higher grades and in high school, and are therefore designed to be applicable to children aged through 17 Approximately 1000 children were tested at each half-year, it was administered routinely to all 8B and 10B pupils after 1941–42, in Chicago schools While this means that the norms are not truly national, they do represent the school population of one of our largest cities and provide useful norms, it would still be desirable to have national norms, but even more important is the accumulation of local norms by other school systems, colleges, and organizations using the tests The battery consists of 11 tests, selected from the 60 tests tried out experimentally on 1154 pupils and subjected to factor analysis, and a second experimental battery of 21 tests tried out on 437 subjects and factorially analyzed These 11 tests measure six primary mental abilities, named Verbal Meaning (V), Space (S) Number (N), Memory (M), Word Fluency (W), and Reasoning (R) These are measured by tests such as vocabulary and opposites (V), flags and cards (S), addition and multiplication (N), and letter grouping (R) Two tests are used to measure each of the six abilities except memory, tested by one test, they are arranged in booklets which can be administered in two school periods Each test is accurately timed, with a practice exercise preceding it They can be scored by hand or by machine, perforated stencils being provided for the former

Evaluation

The success of the Thurstones in constructing a practical battery of tests of primary mental abilities is obviously an important question An easily administered and scored, reliable, and valid battery would represent a major advance in aptitude testing, as it would make possible the measurement of a number of aptitudes which are widely used and which are of varying importance in different types of activities Recognition of the importance of this possibility is shown by the fact that, although Thurstone's experimental tests were published in 1938 and the definitive battery only in 1941, there have appeared, within the years since then,

almost a score of studies of their reliability and validity The short forms should be subjected to even closer scrutiny

Influence on Current Test Construction

The influence of Thurstone's factorial analyses of mental abilities has not been limited to these attempts to validate his tests it has manifested itself in verbal and quantitative scores of the American Council on Education Psychological Examinations which he developed (see pages 114 to 124), in the performance and verbal I Q 's of the Wechsler-Bellevue (see pages 142 to 146) and of the California Test of Mental Maturity, in the Arithmetic Reasoning, Verbal Comprehension, and other similar tests of special mental abilities used in the Engineering and Physical Sciences Aptitude Test (p 341), in the Navy's Basic Classification Test Battery (740), in the United States Employment Service's experimental test batteries (224), in the Psychological Corporation's Differential Aptitude Tests (p 368), and in the Aviation Cadet Classification Tests (214) of the Army Air Forces Test batteries such as the last four yield no I Q 's but, instead, yield part scores which, in a given selection program, are weighted according to their differential predictive value, in accordance with a concept of constellations of abilities needed in various occupations rather than of general ability required in varying amounts in different occupations We have seen that the use of quantitative and verbal scores is still somewhat problematical in the case of the A C E tests, and even more so in those of the California and Wechsler-Bellevue Tests The Engineering and Physical Sciences Aptitude Test is as yet virtually untried in this respect (see pages 341 to 342), and the USES tests, not yet released for general use, have been validated only in a preliminary way on small groups The Aviation Psychology Program (214) used tests of this type to good effect, as demonstrated by the correlations in Table 11, which indicate the differential prognostic value of some of the factorial-type tests for pilot and navigator training Multiple

TABLE 11

COMPARATIVE VALIDITIES OF FACTOR-TYPE TESTS FOR AIRCREW TRAINING

Test	*r Pilot Training*	*r Navigator Training*
Reading Comprehension	19	32
Arithmetic Reasoning	09	45
Numerical Operations	04	26
Mechanical Principles	32	13
Number of Cases	300 to 1,500	8,100 to 10,500

correlations for batteries which also include tests of other types were in the 60's

In view of the widespread influence of Thurstone's work, and the important role which it is playing in shaping the intelligence test construction work now being done, it seems essential to discuss in some detail the practical work which has so far been done with the Primary Mental Abilities Tests

Studies of the Tests as Such Traxler (856) ascertained that the *reliabilities* of the original Primary Mental Abilities Tests were high, judging by both the split half and the retest techniques, but attributed this to the importance of speed in all of the tests Results for 673 freshmen at the University of Chicago were analyzed by Stalnaker (714), also in order to evaluate the adequacy of the standardization of the adult tests He reported that the tests used to measure a given factor had intercorrelations of 20 to 79, the mean being 49 Goodman (297) reported slightly lower coefficients These seem rather low, but the intercorrelations of tests not used to measure the same factor range from — 17 to 49, most of them being under 20 More serious than this, perhaps, is the fact that the items were found not to be in the order of difficulty, and that some items were ineffective His conclusion was that the tests were not yet ready for use with individuals

Adkins and Kuder (8) administered the original Primary Mental Abilities Tests and the Kuder Preference Record to more than 500 freshmen at the University of Chicago, and found relatively little overlapping between the two sets of measures What overlapping there was seemed reasonable in view of the nature of the tests in question Shanner (711) reported a study made at the secondary school level at about the same time He concluded, from evidence generally similar to Stalnaker's, that the tests were reliable and had sufficiently low intercorrelations to indicate independence of the traits measured Although concluding that the tests need more research for refinement and interpretation, he stated that they are a valuable addition to the field of aptitude testing Issue was taken with this conclusion by Crawford (178), who presented the test intercorrelations by frequencies rather than averaging them and concluded that they were not sufficiently independent He also pointed out that the correlations between PMA tests and Co-operative Achievement Tests were low, and concluded that the tests do not have demonstrated diagnostic value Fortunately, more satisfactory evidence is now available, to take the issue out of controversy and into the realm of fact

Applications to Education and Vocations The experimental edition of the PMA Tests was given to 501 University of Chicago freshmen by Shanner and Kuder (712), together with a number of other tests, and correlated with grades on comprehensive examinations taken to secure exemption from freshman courses Results are presented in Table 12

TABLE 12

CORRELATION BETWEEN TESTS OF GENERAL, SPECIAL, AND PRIMARY MENTAL ABILITIES AND FRESHMAN EXAMINATION GRADES, UNIVERSITY OF CHICAGO

Test	*Biological Sciences*	*Humanities*	*Physical Sciences*	*Social Sciences*	*Average Exam Grades*
A C E Psychol Exam	48	48	48	57	52
Physical Sciences Apt	—	—	65	—	52
Social Sciences Apt	—	—	—	65	575
PMA Perception	08	13	17	135	12
Number	21	265	27	30	31
Verbal	38	47	38	435	415
Spatial	225	07	14	13	18
Memory	145	13	18	16	20
Induction	22	03	25	20	23
Deduction	42	19	485	43	38
Multiple R	50	54	56	57	—

It can be seen from Table 12 that the two especially constructed aptitude tests yield the highest validities for the appropriate subjects, and validities at least as high as any other test for average grades It would seem probable, in view of the multiple correlations between PMA tests and subject grades, that these tests would predict average grades about as well as the A C E and the special aptitude tests, were it not for the tendency of multiple correlation coefficients to shrink For special subjects these last have the advantage of being based on job analysis and of being basically miniature situation tests, the presumably greater versatility of the PMA tests makes them more desirable for selection in institutions which do not have large test construction staffs and for general vocational and educational counseling When the PMA Tests are compared with the A C E , it is notable that no single PMA factor is as good a predictor as the test of general scholastic aptitude (although the verbal and deduction factors do about as well for certain courses), and that the multiple correlations between PMA Tests and grades in specific courses, while generally higher than those of the A C E , are not usually sufficiently greater to justify the additional time and effort in test administration and scoring

In a study by Yum (951), also of University of Chicago freshmen, somewhat less promising results were obtained He computed one relationship

not reported by Shanner and Kuder, namely a multiple correlation between PMA Tests and semester average More important still, he used actual grades for students taking the courses, rather than grades based on examinations taken to obtain exemption from the courses. The correlation was .42, which is considerably lower than that obtained by Shanner and Kuder for any single subject, and lower than their multiple correlation would presumably have been had it been computed

Ellison and Edgerton (239) used the experimental tests first published by Thurstone with 49 liberal arts students at the Ohio State University Only the Verbal and Memory Tests had moderately high correlations with point-hour averages (.44 and .31 respectively), but the multiple correlation for weighted scores was .64 The results for grades in specific courses were most promising for Verbal, Spatial, and Deductive Tests gave better predictions of English grades than did the Ohio State Psychological Examination (.75, .44, and .44 as opposed to .42), the Verbal Test predicted Science grades better than the general examination (.68 vs .42), and similar results were reported for foreign language grades and for psychology grades The numbers in each case were, however, between 25 and 30, and the results seem almost too good

Most helpful is a series of studies conducted at the Pennsylvania State College under the direction of Robert G Bernreuter Ball (40) administered the older Thurstone battery to 147 freshmen women and 159 men in the liberal arts college The correlations with semester point average ranged from .04 for the Spatial Tests to .35 for the Verbal The multiple correlation for Memory, Number, Verbal, Induction, and Deduction Tests and semester point average was .46, which is no better than what one would expect from a much briefer scholastic aptitude test Some of the tests of specific factors correlated substantially with appropriate college marks, the coefficient for Number and Mathematics being .41, and Verbal and English Composition .40 The Verbal Tests, however, tended to have moderately high correlations (.20 to .40) with all courses Hessemer (369) analyzed PMA Test scores for 147 freshmen women, using first semester point average and grades in inorganic chemistry as her criterion The Verbal Tests were again the best predictor, with a correlation of .44 with semester point average, Deduction followed closely with one of .40 There were no satisfactory correlations, however, with chemistry grades, that for the Verbal Tests being .13, and the two highest being −.25 for the Spatial Tests and .18 for the Deduction Tests, the irreconcilability of which relationships suggests their chance nature

Bernreuter and Goodman (88) obtained data for 170 freshmen engineers In this instance the correlations between PMA Tests and semester point average ranged from 04 (Perceptual) to 38 (Deduction), the multiple correlation being 51 for Number, Verbal, Space, Induction and Deduction or Reasoning Tests Again the Verbal Tests yielded significant correlations with all courses (except Drawing), the correlation between Verbal Test and English Composition grades was 44, and that between Number and Mathematics grades was also 44 Unfortunately this study, like the others just summarized, provides no validity data for tests of general intelligence, which might enable one to decide whether or not the extra time required by the PMA battery is justified by higher validities This defect is remedied by another Penn State study by Tredick (869) who tested 113 freshmen women students of home economics with the PMA battery, the Otis, and several other tests The results, shown in Table 13, are in line with the trend of those so far reported, in that the Verbal Tests tend to give moderately high predictions of grades in all courses and especially in English (55), the Number Tests have a substantial correlation with chemistry grades (46), and Induction and Deduction are also good predictors Most interesting, perhaps, is the fact that the multiple correlation coefficient of 61 for four PMA Tests (NVID) and semester point average is substantially higher than that of 53 between the Otis and the same criterion, but the R was apparently not corrected for the shrinkage which usually takes place with a second group

It is interesting to note, in passing, the correlations between PMA Tests and the tests of general and special aptitudes used by Tredick All of the former have moderate or high correlations with the Otis (29 to 68), only the coefficients for the Number and Memory Tests being below 40 The perceptual factor is important in the Otis ($r = 53$) (presumably because of the emphasis on speed), the Minnesota Vocational Test for Clerical Workers (57, 51) and the Minnesota Spatial Relations Test (55), but much less so in the Minnesota Paper Form Board (39) The number factor is highly correlated not only with the Minnesota Clerical Numbers (59), but also with the Names (58) Test The verbal factor is very important in the Otis (68), and of moderate importance in the Clerical Names (40) and Art Judgment Tests (39) The spatial factor plays a moderately important part in all of the tests in the study except, interestingly, in the Art Judgment Test, where its role is of only slight importance (20), it is most closely correlated with the Minnesota Spatial

TABLE 13

CORRELATIONS BETWEEN PMA TESTS, OTHER TESTS, AND GRADES FOR 113 HOME ECONOMICS FRESHMEN (TREDICK)

	Sem Pt Average	*Art*	*Engl Comp*	*Chemistry*	*P*	*N*	*V*	*S*	*M*	*I*	*D*
PMA Tests Combined NVID	61										
P	28	15	19	20							
N	41	11	22	46							
V	51	24	55	28							
S	28	25	10	23							
M	20	− 02	08	25							
I	40	26	19	37							
D	42	21	21	43							
Otis	53	17	54	37	53	33	68	40	29	60	61
Minn Paper Form Board	31	24	13	24	39	21	24	37	06	48	45
Minn Clerical Names	24	07	07	26	57	58	40	41	24	44	46
Minn Clerical Numbers	36	08	27	31	51	59	06	36	20	28	28
Minn Mechanical Assembly	11	17	− 01	16	23	− 12	11	34	07	26	30
Minn Spatial Relations	23	20	02	22	55	15	16	49	06	47	33
Meier Seashore Art Judgment	23	20	29	03	33	11	39	20	18	23	15

Relations Test, but only to the extent of .49, and its relationship to the Minnesota Paper Form Board is no closer than to the other non-spatial tests (.37 as contrasted, e g , with .41 and .36 for Clerical Names and Numbers) This suggests that the so-called spatial factor measured by the PMA Tests may be more general than strictly spatial The memory factor is moderately correlated only with the Otis (.29), other coefficients are about .20 or below Induction plays important parts in the Otis (.60) and in the spatial tests (.48 and .47), is moderately important in Clerical Names (.44), and of some importance in the other tests used The deduction or reasoning factor plays a similar role, but is somewhat less important in

the Minnesota Spatial Relations Test than in the Paper Form Board (r's of 33 and 45 respectively)

Goodman (297) reviewed the work done by other investigators at Penn State, and reported further research of his own with engineering freshmen The correlations between PMA Tests and first year semester point averages ranged from 08 (P) to 34 (V) and 36 (D) The Number factor, which one might logically expect to yield one of the highest r's with engineering grades, had a correlation of only 36 and the spatial factor only 18 This tends to support the conclusion drawn earlier from work with the A C E part scores, to the effect that verbal and "general" intelligence tests are at least as effective predictors of success in technical courses as are more quantitative tests Goodman also obtained the intercorrelations between specific tests in the PMA battery, and between these tests (as contrasted with combinations of tests which measured specific factors) and the criterion This analysis showed that the intercorrelations of tests measuring the same factors ranged from 01 to 72, with a median of 33, which suggests that the measurement of specific or primary factors still leaves much to be desired, it also revealed that some of the specific tests had higher correlations with the criterion than did the factor scores to which they contributed This last finding is not surprising, as a test of mixed factors might predict success in a task involving some of those same factors better than a score representing more adequately one "pure" factor which is only one contributor to success in the activity in question

A few other studies which have been reported show results similar to those just reviewed Stuit and associates (787) administered the PMA Tests to students in engineering, medical, and journalism schools, and reported characteristic profiles Engineers were high on S and D, low on V and M, journalists were high on P, N, and V, low on M and D, medical students were high on P and I This suggests that the battery should be useful in guiding students into curricula in which their abilities resemble in type those of the majority of students More work should be done along these lines, as the differential use of the tests should be one of their principal contributions However, this type of standardization has merely been begun

Perhaps the nearest thing to validation in terms of vocational criteria has been carried out by Harrell and Faubion (338), who administered the experimental PMA Tests to 105 men in aviation maintenance courses in an Air Forces Technical School The multiple correlation of

Verbal, Spatial, Induction, and Deduction Tests with average grades was 63, which contrasted with a correlation of 45 between Army Alpha scores and grades The Number Tests correlated most highly with grades in shop mathematics (37, contrasted with 31 for Army Alpha and 46 for the combined PMA Tests), the Verbal Tests correlated most highly with grades in Electricity (51, compared to 47 for Army Alpha and 57 for the combined PMA Tests), and the Deduction or Reasoning Tests predicted grades in blue-print reading and mechanical drawing most effectively (54, compared with 30 for Army Alpha, 36 for the Spatial Tests, and 60 for the combined PMA battery)

Use of the PMA Tests in Vocational Counseling and Selection The studies reviewed in the preceding pages make it clear that the long forms of the Thurstone Tests of Primary Mental Abilities, while sufficiently perfected to make possible important research into the nature and organization of human abilities, still need to be improved before they become a practical instrument for use in guidance and selection The defects in the tests have been summarized by Crawford and Burnham (180 213) The measures of specific factors are still somewhat impure, as shown by the moderate rather than high intercorrelations of tests used to measure a given factor Speed plays too important a part in all the tests The relationships between specific factors and other tests or criteria with which they might be expected to be related are often low enough to make one question the adequacy of the measurement of the factor (e g , the spatial factor) On the other hand, there are a number of findings which are extremely encouraging Among these are the generally higher multiple correlations between PMA Tests and criteria than among general intelligence tests and criteria, which suggest that in selection work especially it will be advisable to use this more refined type of measure, to obtain differential occupational weights, and to score accordingly In time, accumulated data may make these differential weights useful in counseling that is, the score for an improved spatial factor might be multiplied by 5, that for the number factor by 4, that for the verbal factor by 2, etc , in order to compare the promise of a counselee with that of others who have entered technical occupations, whereas the same scores would be multiplied by weights of 1, 5, and 4 in order to compare his promise with that of men in accounting occupations This technique was applied to potential pilots, navigators, and bombardiers in the Army Air Forces with considerable success, is being experimented with by the United States Employment Service, and may

become possible with the PMA Tests as they are improved and occupational norms are accumulated In the meantime, it should be remembered that these tests are still a promising device for research rather than a practical tool for counselors or personnel managers, and that the short forms are as yet untested

The Wechsler-Bellevue Scales of Mental Ability (Revised manual Williams and Wilkins, 1943 Test materials Psychological Corporation)

The publication, in 1939, of the Wechsler-Bellevue Scale of Mental Ability as an individual intelligence test designed for use with adults rather than with children immediately focussed the attention of the more clinically minded psychologists and counselors on this instrument, even when the nature of the counseling problem was largely vocational and educational The aura surrounding individual testing, as opposed to the supposedly less sensitive measurements obtained from group tests, alone was a sufficient cause of such interest in the Wechsler-Bellevue To this appeal was added, however, that of a test which yields two types of scores, one based on verbal and one on performance items The fact that the scale was developed in a mental hospital, primarily for the diagnosis of mental defects and mental impairment in adolescents and adults, and that all the original material on the test was directed toward these uses (914), resulted only in greater confidence on the part of the clinically minded who proceeded to use the scale in vocational and educational guidance Because of their widespread use in guidance centers some aspects of the Wechsler-Bellevue Scales are considered here, more as a caution to users than as a guide to use in vocational counseling

The question of the clinical usefulness of the scales is clearly quite independent of the question of its usefulness in vocational counseling and selection When considering the use of such a test in vocational guidance and personnel work, three questions are relevant First, what advantages, if any, does an individually administered test of mental ability have over a group-administered test in vocational guidance or selection? Secondly, how good is the instrument as a test of general mental ability? Thirdly, what evidence is there concerning the occupational significance of total and part scores, particularly the latter? Each of these questions will be dealt with briefly in the following paragraphs

Individual vs *Group Tests* The relative advantages and disadvantages of group and individual, performance and paper-and-pencil tests

have been discussed in Chapter 4 But for the sake of convenience a few especially pertinent points should be made here Tests designed for group administration can also be administered individually, and therefore have the advantage of being more flexible in their use On the other hand, they are generally paper-and-pencil tests, which do not have the flexibility that orally administered individual scales such as the Wechsler-Bellevue possess In the former, the examinee reads and answers questions by himself without the examiner being able to judge his reactions by anything more than expression and gestures, and with no possibility of modifying the questions to suit the background of the subject In the latter, the administrative procedure is more conversational, and therefore the examiner has much more opportunity to judge the reactions of the subject and to modify procedures in such a way as to be completely fair In clinical work the desirability of the latter type of technique is obvious, for then one is working with cases whose background or condition is unusual in some respects and it is important that the test situation permit the examiner to observe these abnormalities and to modify the test procedure accordingly in some instances and to note them for diagnostic use in others But in vocational and educational counseling or selection the examiner is dealing with persons whose condition is approximately normal and whose background is such as to make standardized techniques appropriate For each normal counselee or employment applicant there is a suitable group test of mental ability, developed for use with and standardized on subjects such as he modification of test procedures is therefore generally unnecessary if the examiner has background data on his subjects and chooses his tests well Furthermore, the normality of the examinee means that the purpose of the test is to get an overall measure of mental ability, not to study peculiarities of mental functioning For this reason also the group test, which provides a suitable series of standardized tasks and obtains a measure of performance on those tasks, yields all of the types of data which can legitimately be expected from intelligence testing for vocational or educational purposes

The Wechsler-Bellevue as an Intelligence Test Studies of the Wechsler-Bellevue published prior to 1945 have been summarized by Rabin (618) and by Watson (912) The trends revealed in these summaries are for the Wechsler-Bellevue scores and Revised Stanford-Binet to correlate from .78 to .93 when the groups are heterogeneous in age or mental ability, and about .62 when they are more homogeneous

(e g college freshmen) The verbal scale is uniformly more highly correlated with the Revised Stanford-Binet than is the performance scale The correlations with group tests are, as has generally been the case with individual tests, lower than those with other individual tests for Army Alpha a coefficient of 74, for the Otis SA 425, and for the A C E 48 and 53 are reported Wechsler-Bellevue I Q 's of superior individuals were found to be lower than those obtained on the Revised Stanford-Binet, while persons of little mental ability made higher scores on the Wechsler than on the Binet Scales, because the Wechsler has a smaller standard deviation Rabin and Watson also deal with the clinical significance of part scores, but that topic is not relevant to our purposes From the trends reported above one can conclude that the results of the Wechsler-Bellevue Scales agree with the results of other intelligence tests as well as can be expected

Occupational Significance of Total and Part Scores From the point of view of the vocational psychologist, counselor, and personnel manager, the crucial question concerning this or any other intelligence test is what evidence is there to help me interpret the test scores in terms of prospects of success in various types of work? The answer, for the Wechsler-Bellevue Scales, is practically none Neither Rabin nor Watson located any studies of the occupational significance of Wechsler-Bellevue scores, and the writer has located only one published prior to 1947, in which Altus and Mahler (15) reported significant differences in the Wechsler Mental Ability Scale (Form B) verbal scores of 2476 Army illiterates who had been employed in skilled or semiskilled occupations, on the one hand, or in unskilled occupations on the other One can, of course, use the total or verbal scores in a general way, by analogy A person who has very superior intelligence on these scales would also have very superior intelligence on Army Alpha, and such people, we know from research with the latter test, tend to succeed in the higher professional and managerial occupations, similarly for dull, normal, average, and other levels But the possibility of such interpretations does not constitute a special advantage of the Wechsler-Bellevue in vocational and educational guidance It is, rather, a means of salvaging and making useful the results of a test which would otherwise be useless in vocational guidance and selection There are other tests of mental ability whose vocational significance is based on more direct evidence they are therefore subject to less error in interpretation

For the part scores, or verbal and performance I Q 's, the answer

concerning the vocational and educational significance of the scales is even less equivocal Neither Rabin nor Watson mention the occupational significance of verbal and performance I Q 's, although Rabin cites one study (16) of the relationship between total and part scores and achievement in college In this investigation Anderson and his associates reported a correlation of 41 between the full scale and the first semester grades of 112 college women, while the Verbal and Performance I Q 's yielded correlations of 50 and 19 respectively These compared with correlations for 1941 A C E total, linguistic, and quantitative scores of 54, 54, and 39 (the data for the 1940 form of the A C E were 48, 48, and 36) Obviously, the Wechsler-Bellevue Performance I Q is of no value in predicting success in the first semester of a liberal arts college, and the performance items lower the validity of the verbal items in the total score The Verbal I Q itself is no more adequate a predictor of success in the liberal arts than is a group test of intelligence such as the A C E

With such a paucity of evidence the use of the Verbal and Performance I Q 's in the differential diagnosis of vocational and educational aptitudes is clearly unwarranted To reason by analogy and interpret Wechsler-Bellevue scores as though they were synonymous with linguistic and quantitative scores on the American Council Psychological Examination or with primary mental abilities scores on the Thurstone tests is also unwarranted, although this seems to have become a rather widespread practice among psychometrists and counselors. It is true that Balinsky's factor analysis (39) isolated verbal and performance factors, the former consisting, at age 25–29, of digit-symbol, comprehension, and information items, and the latter of "spatial" items such as picture completion, object assembly, and block design But Anderson and others (16) have shown that, although there is a moderate correlation between Wechsler Verbal and A C E Linguistic scores ($r = .49$ or .50), the relationship between Performance and Quantitative scores is too low ($r = .31$ or .39) for interpretation of one in terms of the other to be justifiable No such data are as yet available for the Wechsler and PMA Tests And we have seen that differential educational diagnosis on the basis of either A C E or PMA Test part-scores is still in the experimental stages

Use of the Wechsler-Bellevue Scales in Counseling and Selection As the Wechsler-Bellevue Scales are used in more and more studies evidence upon which to base judgment concerning the vocational and educational significance of part scores will presumably be forthcoming In the mean-

time the objective psychologist, counselor, or personnel officer can only recognize that the use of anything more than the total or verbal score as a rough index of the educational and occupational level which the person in question may attain is unwarranted, and that, for most persons, this can be done at least as well and more economically by means of paper-and-pencil tests

CHAPTER VII

PROFICIENCY

PROMISE AND PROFICIENCY

IN COUNSELING young people concerning the choice of careers one is generally concerned with *promise*, that is, with prospects of success in a field in which the youth has as yet had no substantial training or experience In selecting employees, on the other hand, the concern is more likely to be with *proficiency*, that is, with present ability to perform the tasks involved in a given job Proficiency, achievement, or trade tests are therefore generally thought of as instruments for the selection of personnel or for the evaluation of the outcome of training, whether in school or on the job However, past achievement is often one of the best indices of future accomplishment, so that achievement tests can frequently be used as tests of aptitude for related types of activity

The difference between an aptitude and an achievement test therefore lies more in its use than in its content An achievement or proficiency test is used to ascertain what and how much has been learned or how well a task can be performed the focus is on evaluation of the past without reference to the future, except for the implicit assumption that acquired skills and knowledge will be useful in their own right in the future A test of achievement in arithmetic is therefore a measure of mastery of the essential processes of arithmetic and of ability to make certain types of computations A measure of proficiency in typing is an index of ability to copy typewritten material with speed and accuracy and therefore of ability to perform certain types of clerical duties to an employer's satisfaction An aptitude test is used to judge the speed and ease with which skills and knowledge, that is, proficiency, will be acquired But, obviously, proficiency in a given task may be an index of promise in a related task, and knowledge of certain types of facts may be indicative of facility for the learning of other types of facts

Therefore a test of arithmetic achievement may be a good index of aptitude for algebra or for engineering, a test of typing proficiency may

be a good measure of aptitude for stenography, and a test of information concerning recent developments in science may be a good predictor of success in medical training Each such relationship is of course strictly hypothetical until experimentally checked and found to be true, for even a good achievement test cannot be assumed to be a good aptitude test until it has been validated in the same manner as any other aptitude test Achievement in arithmetic may prove to predict success in algebra, but have no relationship to engineering grades one cannot take the relationship for granted, since what may seem like perfectly legitimate assumptions in the field of prediction often prove unwarranted An achievement test (or test of any type) can be used as an aptitude test only when there is a known relationship between the performance tested and the performance in which success is to be predicted This is the essence of aptitude testing, the understanding of which takes all the mystery out of the subject As it becomes more generally realized that aptitude testing is nothing more than the prediction of success in one performance by means of a measure of success in another performance known to be related to it, people in need of guidance will have more reasonable expectations from tests, business and industrial men will be more inclined to see their possibilities and limitations, and professional users of tests will have more freedom to make legitimate use of them

Educational Achievement Tests

Educational achievement tests are of interest to us here only as indices of promise in vocational activities Treatises of their use in evaluating the results of instruction, as measures of educational progress, and related topics are numerous (310,474,650) unfortunately there has been less study of their use in predicting educational success and still less of their value in predicting vocational success

In the prediction of educational success, educational achievement tests have been effectively used in the admission programs of colleges and professional schools In most investigations they have been tried in combination with tests of scholastic aptitude and with high school averages, in order to determine the relative value of each type of predictor In one such study at the University of Minnesota (930) it was found that high school rank was the best single predictor of sophomore achievement, but that a combination of the three types of indices was better than any single index They may be similarly used by counselors in guiding students concerning the choice of college or professional school, the coun-

selee's standing on a test in comparison with typical candidates for admission being used as an index of the possible wisdom of the choice

It would be desirable also to have data which would make it possible to counsel concerning the wisdom of the choice of a major field of study such as premedical, engineering, business, and related courses but unfortunately the data which are needed for such applications of achievement test results are available for only a few institutions Although the assumption that a *weakness* in science after high school should be taken as a negative indication for a college major in science has some justification, it should not be concluded that the *lack* of a given high school subject will mean a weakness in a related college subject, for too many studies of the importance of high school prerequisites in college admissions (110) have demonstrated that one may do superior college work regardless of background in specific high school subjects On the other hand, it seems likely that the quality of work done in a high school subject, whether measured by the grade obtained or by score on an achievement test, will, other things being equal, be indicative of the quality of work that will be done in a college course in the same subject The question is, what is known concerning the actual predictive value of achievement tests? This question will be examined in connection with the specific tests discussed in the paragraphs which follow In brief, experience has shown that achievement tests not only yield predictions of college averages which are about as good as those provided by intelligence tests, but also give better differential predictions of success in specific subjects than do intelligence tests (117,701)

The Iowa Placement Examinations (University of Iowa, 1924, 1930, 1941)

These were among the first educational achievement tests which were constructed, under Stoddard's supervision, to cover the major subject-matter fields for purposes of differential prediction in college First published in the mid-twenties, they have been widely used and are among the best constructed and most thoroughly understood tests of their type, hence their treatment here

Applicability, Content, Administration, Scoring, and Norms There are two series, one designed to measure achievement and assuming a year of high school work in the subject, the other designed as a measure of aptitude for the same subjects Both are designed for placement in college classes Fields covered are Chemistry, Physics Mathematics, English, French, and Spanish The training or achievement series has been the

more widely used, and has been generally the more valid The tests require forty minutes each for the administration, and scoring is by means of a convenient stencil Normative groups are large, consisting of more than 10,000 students in nine colleges

Standardization and Initial Validation In constructing subject-matter tests the attempt is generally made to obtain what Mosier (548) calls *validity by definition* by having experts outline the content of the field to be covered by the test and construct items which they consider to be representative of that content, these outlines and items are then checked by other experts in the same field, in order to insure representative judgment Textbooks and courses of study were analyzed in the development of outlines The tests were then correlated with first semester marks in the nine co operating colleges, the correlations ranging from 26 to 95, the mean for the Training Series being 60 and that for the Aptitude Series 50 (759) Stoddard reported that the Iowa Placement Examinations gave better predictions of grades in specific subjects than did either high school grades or intelligence test scores, and he and Hammond (326) found that the combined achievement scores had more predictive value than an intelligence test, although he also found that a single intelligence test gives a better prediction of average college marks than does a single achievement test

Reliability As might be expected in the case of carefully constructed achievement tests, the reliabilities are high they ranged from 87 to 92 (759)

Validity Validation of the tests subsequent to their development was pursued most intensively at the University of Iowa in the late '20's and at the University of Minnesota ten years later Hammond and Stoddard (326) used the tests in a number of engineering colleges, with results comparable to those first obtained by Stoddard The extremely high and low scores were found to be especially useful in singling out students who were most likely to succeed and fail, respectively For example, of the 100 highest and 100 lowest scoring students on the mathematics achievement test, only seven of the former and as many as 61 of the latter failed the first semester course in mathematics Working with engineering freshmen at Minnesota, Northby (569) found correlations of 55 and 70 between the same test and honor-point ratio for two different classes In all the groups studied by Hammond and Stoddard the proportion of failing students in the top quarter of the placement examinations was less than 10 percent, while from 28 to 58 percent of those in the lowest quarter

failed the first semester's work Segel summarized research with these tests in 1934, finding a median correlation of .40 between the English Training Test and college English marks and one of .54 between the Chemistry Training Test and college marks in Chemistry (701)

Use of the Iowa Tests in Counseling and Selection The data summarized above make it clear that the Iowa Placement Tests, especially the achievement tests, can be used as indices of differential achievement in college Students or prospective students who show strength in a given subject-matter test are more likely than not to make good grades in that field, whereas those who make low scores despite appropriate preparation are not likely to make good grades in courses in that subject Those whose average on the battery of tests is high are likely to make high grades in their college work taken as a whole The tests can therefore be used in counseling students concerning choice of college and concerning choice of major field, they can also be used in selecting qualified students for admission to college and to departments or professional schools It should be remembered, however, that the tests are not replaced annually by new forms as in the case of the Co-operative Test Service tests, to be described below For this reason the Iowa tests are valuable for the knowledge they provide concerning the predictive value of such tests, but are now of less practical use than some of those developed by active test construction organizations

The Co-operative Achievement Tests (Co operative Test Service, periodically)

The Co-operative Test Service began the publication of annual editions of achievement tests in the major school subjects early in the 1930's, sponsored by the American Council on Education and operated under the leadership of Ben D Wood and John C Flanagan during the first decade of its existence It is now part of the Educational Testing Service

Applicability, Content, Administration, Scoring, and Norms Each test is designed for use at a specified educational level, which may include a range of as many as three or four years The content is kept up to date by the periodic publication of new editions, but earlier editions are also available and are generally usable for several years (an important point, as examination of the content of some well known social studies tests of pre-war vintage will reveal) Norms are provided for large groups of students, and are made national and kept up to date by the large-scale testing programs in which the annual editions are used The content

varies with the field covered by the test and with the level for which it is designed, the method of construction (discussed below) providing for adequate coverage

Of special interest in vocational and educational guidance and selection are the Co-operative Survey Tests (Natural Sciences, Social Studies, and Mathematics), the Co-operative Test of Recent Social and Scientific Developments, designed for use with high school juniors and seniors, the Co-operative General Culture Test (History and Social Studies, Literature, and Fine Arts), and the Co-operative Contemporary Affairs Test, designed for use with college sophomores but applicable to other persons of college caliber and age These tests have the advantage of providing not only comparisons of the achievement of the person being studied with that of other persons with similar backgrounds, but also a picture of the relative strengths and weaknesses of the counselee in the subjects tested The Survey Tests, being based on the content of high school courses, are useful in counseling high school seniors and entering college freshmen concerning the choice of college majors, the other tests, less closely anchored to specific courses, are useful in helping non-science majors understand their special strengths and can be used as something of an interest test, for they reflect to a considerable extent the subjects in which the student has been interested and upon which he has kept informed

Administration is simple, and scoring can be done either by hand or by IBM test-scoring machines, in either case, the use of special answer sheets makes for economy of materials and ease of scoring

Standardization and Initial Validation As in the case of the Iowa tests, the Co-operative Achievement Tests are developed by subject matter experts who work with test technicians, test outlines are based on analyses of courses of study and textbooks, and items are checked by both types of experts The first type of validity achieved is, therefore, validity by definition Further validation is occasionally carried out by correlating test scores with high school or college grades, these correlations have generally been moderately high (.30 to .50) for appropriate subjects (24)

Reliability The reliability coefficients vary slightly with the test and form, but have generally been .90 or higher, as one would expect in the case of subject-matter tests constructed by experienced technicians

Validity It has already been stated that the validities of the Co-operative Achievement Tests for the prediction of grades in related subjects range from about .30 to about .50 When scores made on a battery of

achievement tests such as the Co-operative General Culture Test are combined, higher correlations are reported In one study (24) the validity coefficient for the latter test and average grades for the first two years in college was 53

More striking still are the mean General Culture scores made by students in different major fields, which show that students of journalism, religion, and law made above average total scores, probably reflecting their broader interests, whereas engineers have apparently a much more restricted range of interests and make significantly low general culture scores More important than pre-occupational differences in total general information are, of course, the differences in patterns of scores on the various subtests Analyses (679) of these show that students who later became medical students had made generally high scores as freshmen Dentistry students had excelled in mathematics and science but not in other areas Journalism students reversed this pattern Library Science students were high in English but mediocre in other fields Business students were characteristically high in mathematics but low in English

It would be desirable to have data showing the relationship between patterns of achievement on tests such as the Survey and General Culture Tests, and choice, achievement, and satisfaction in different types of work One would expect, for example, that social workers would be persons who, in college, made their highest scores on tests of the social studies, and that successful engineers are those who, on entering college, showed special strength on tests of achievement in natural sciences But no data such as these have come to the writer's attention

Use of the Co-operative Tests in Counseling and Selection In view of the moderately high relationship between scores on these subject-matter achievement tests and grades in appropriate courses, they may well be used in helping students evaluate their prospects of success in various major fields in high school and college, in placing students in sections for which their background qualifies them, and in selecting students for courses of training which emphasize mastery at a higher level, of the same type of subject matter as that covered by the test There are as yet no direct, objective data to justify counseling concerning the choice of an occupation on the basis of educational achievement test scores, but insofar as achievement on a test is related to grades in a professional or vocational school, and grades in such a school are related to entry into or success in the occupation for which it prepares, it should be safe to

deduce that some educational achievement tests do have at least indirect predictive value for some occupations

The Tests of General Educational Development (Science Research Associates, 1946)

The Tests of General Educational Development were constructed for the United States Armed Forces Institute under the direction of E F Lindquist, another series under the same title was developed by Lindquist at the University of Iowa Both series are obtainable from Science Research Associates As the USAFI series has the most comprehensive norms and has been most widely used with returned servicemen it will be discussed here

Applicability, Content, Administration, Scoring, and Norms The GED Tests were designed for use at high school and college levels, a separate battery being designed for each level The high school battery covers five areas correctness and effectiveness of expression, interpretation of reading materials in the social studies, natural sciences, and literature, and general mathematical ability At the college level there are four tests, mathematical ability not being covered The objective is to measure understanding rather than factual knowledge The tests are power tests, with two hours allowed for each test IBM answer sheets make possible stencil or machine scoring Norms are available for six geographical regions (an advance over national norms) and for the country as a whole, and for college students in three types of institutions classified according to freshman mental level

Standardization and Initial Validation The procedures used in developing the GED Tests were similar to those used in the construction of other achievement examinations, with the exception that an attempt was made to measure understanding rather than factual knowledge in view of the lapse of time since many service-men had attended school This trend is a wholesome one in achievement testing, in view of the common tendency to overemphasize factual knowledge, it should not result, however, in failure to measure the mastery of factual knowledge which constitutes the basic tool of many subjects

Reliability In view of the attempt the test authors made to measure understanding rather than knowledge of facts, it is perhaps important to note that the reliabilities of the tests are not reported

Validity The validity of the GED Tests has been studied primarily

in relation to the prediction of success in college. Crawford and Burnham (179) administered the tests to 135 freshmen (veterans and non-veterans) at Yale University, finding a correlation of .72 between total scores on the GED Tests and the College Entrance Examination Board Tests, and correlations of .56 and .53, respectively, for each of these tests with first-term freshman grades Correlations between part scores on the GED Tests and freshmen marks ranged from .36 to .51, the former being for the natural sciences and the latter for the expression test Dyer (227) tested a group of 114 Harvard students, about one-half of whom were freshmen and the balance in the three other classes He also found that the total score provided a reasonably good prediction ($r = .46$) of college grades In a third study conducted at the University of Minnesota, Callis and Wrenn (131) obtained a correlation of .72 between total GED score and honor-point ratio The significantly higher figure may be a chance error related to the small number of cases ($N = 56$), or may perhaps be due to a greater range of ability resulting from less stringent initial selection in a state university The authors' comparison of the two suggests the latter Like the other studies, this one suggested that the Expression and Social Studies Tests are the best single tests in the battery for predicting overall success in college

The value of the GED Tests in placing students in advanced courses, one of the principal uses to which the tests were intended to be put, has been ascertained only by Dyer (227) He found that, with curricula and promotions such as those at Harvard, the tests were of no value in the advanced admission of students in either scientific or non-scientific curricula Dyer also reported that patterns of GED scores tended to agree with patterns of interest as shown by field of concentration

Use of the USAFI GED Tests in Counseling and Selection The studies so far published indicate that the GED Tests are scholastic aptitude rather than scholastic achievement tests In view of the authors' attempt to measure understanding rather than factual knowledge this finding should not be surprising They can therefore be used in counseling students concerning the choice of colleges, and in selecting students for admission They have some differential value for science and non-science majors, just as do achievement tests in appropriate subjects, and just as the part scores of scholastic aptitude tests show promise of doing They have, finally, the advantage of not looking like or being labelled as intelligence tests, which may make them more acceptable for use with some candidates for college entrance

Vocational Proficiency Tests

Unlike tests of educational achievement, vocational achievement tests cannot well be used as measures of aptitude in students vocational proficiency tests measure skills or knowledge already acquired, rather than ability to acquire them The degree of skill or knowledge in one occupation cannot often be used to predict the degree of skill which will be acquired in another occupation, even when the latter can be considered a higher level occupation of the same general type skill as a machinist is an inefficient index for the prediction of success in engineering, for training in the one occupation takes as long as training in the other and the varying degrees of aptitudes needed in each can be more easily measured by other types of tests On the other hand, vocational proficiency tests can be used as indices of the prospects of success in a job when dealing with trained candidates for a job Such tests are therefore widely useful in selection, but in counseling only with marginal workers who may need to be encouraged to change their field of endeavor Because they are largely a selection technique, and because companies with good selection devices of their own developing generally prefer to keep them from becoming known to others, there is little published material concerning specific vocational proficiency tests Apart from a few stenographic tests, the trade tests developed by the U S Army, Navy, and Employment Service are the most generally known

The Blackstone Typewriting Test (World Book Co , 1923)

Although one of the first tests developed to measure proficiency in typing, this test is still a standard test in this field Developed for testing students' proficiency in courses, it is also useful with employment applicants

Applicability, Content, Administration, Scoring, and Norms The test can be used at and above the high school level, with persons who have had some training in typing It consists of a typical business letter to be copied by the examinee. It can be given in group form, requiring only three minutes The score is the number of errors and corrections Norms are based on more than 2000 cases with from five to 20 months of instruction

Standardization and Initial Validation Typical business letters were analyzed to determine the average number of strokes per word and various forms were tried with varying time limits and scoring methods The

final forms distinguished clearly between students with differing amounts of training No validation against success on the job was attempted, since it was thought of as an educational achievement test

Reliability The average inter-form reliability was reported as 93 in the manual, the students in question having had twenty months of training

Validity There seems to be a tendency to assume that tests such as this are valid, examination of the content showing that it is a typing test It would still be desirable to know what the relationship is between speed and accuracy of transcription in a relatively artificial test situation and speed and accuracy in a routine work situation

Use of the Blackstone Typing Test in Selection Each organization using a test such as this should empirically determine its own cut-off scores by ascertaining the range of scores on its employees and setting the critical minimum score at a point which eliminates those who are too slow or inaccurate

The Blackstone Stenography Test (World Book Co , 1923)

This test is designed to measure more than ability to take and transcribe dictation, and to include English, office practice, and related abilities

Applicability, Content, Administration, Scoring, and Norms This test also was designed for use at or above the high school level The English test measures knowledge of grammar, punctuation, capitalization, and spelling by means of sentences in which the type of error made is to be indicated, three tests measure proficiency in hyphenating, alphabetizing, and abbreviating, two tests cover knowledge of office practice and business organization, and one test measures ability to take dictation at a fixed rate and to transcribe two letters on the typewriter. This is a group test, but the two letters to be dictated are to be chosen from the manual on the basis of appropriateness to the persons being tested Dictation time ranges from one to three minutes, transcription time is twelve minutes, and other parts require 33 minutes Norms are based on 1000 students with varying amounts of training

Standardization and Initial Validation Correlations of 62 and 79 with efficiency ratings for groups of 37 and 49 stenographers are reported in the manual These seem remarkably high, but the data do not permit evaluation of the adequacy of this phase of the work with the test

Reliability The inter-form reliability reported in the manual is 88 for 1000 subjects

Validity No validity data have been located in the literature, although they would be even more desirable in the case of a test such as this than in the case of the typing test, since it purports to measure more than one phase of stenographic work

Use of the Blackstone Stenography Test in Selection In the case of this test also, local critical minima should be established with the aid of employee evaluation techniques Such minima must, of course, vary with the available supply of workers

The Seashore-Bennett Stenographic Proficiency Tests (Psychological Corporation, 1943)

The Seashore-Bennett tests are a new series of phonographically recorded stenographic proficiency tests, two forms designed for use in employee selection by business firms, the others for use in schools and employment agencies The use of recordings of business letters was resorted to as a method of standardizing the voice and rate of dictation

Applicability, Content, Administration, Scoring, and Norms These tests have, like others in this category, been designed for use at the high school level or above, with persons who have had some training in shorthand and typing They consist of phonographically recorded letters, five letters (four discs) to each form of the test Two letters are short and slow, two are of medium length and average speed, and one is long and rapid Administration requires about fifteen minutes, with another half hour for transcription Complete scripts and reproductions of good and poor transcriptions are provided for use in scoring Norms are not provided, as it was expected that they will vary considerably from company to company, and nation-wide norms could not be collected Distributions of scores for several companies are provided in an article published subsequently to the manual (697)

Standardization and Initial Validation In one sense, this test can depend on internal evidence of validity, for it involves shorthand and transcription It is virtually a life-situation test Preliminary validation studies have been reported, however, showing correlations of 49 and 61, respectively, with supervisors' ratings of general value (combined ratings) and stenographic ability (697)

Reliability When scores on two of the letters were correlated with scores on the other three, the reliability coefficients were 80, 83, and 91

Validity The tests are too new for other validation studies to have been completed and published

Use of the Seashore-Bennett Tests in Selection The availability of alternate forms of these tests makes possible their use both in initial selection and in the evaluation of progress for promotion Local norms for both purposes should be developed, as job requirements vary both within an organization and among organizations It may be found desirable, for example, to start new stenographers in certain departments but not in others, transferring them to these latter positions after promotion tests demonstrate that they have attained the proficiency needed in the more demanding positions

The Elwell-Fowlkes Bookkeeping Test (World Book Co , 1928)

This test was developed for the evaluation of progress in bookkeeping courses, and, secondarily, for judging applicants for positions It covers the first two semesters of bookkeeping

Applicability, Content, Administration, Scoring, and Norms It may be used with high school students and adults who have had some training in bookkeeping Two tests are available for the two semesters of bookkeeping There are nine parts covering theory, journalizing, classification, adjusting entries, closing the ledger, and statements Administration time is about one hour Norms are based on about 250 students in each semester

Standardization and Initial Validation The test covers standard course material and, like most achievement tests, depends upon face validity and care in construction

Reliability Inter form reliability is 82 and 87 for the two levels of the test

Validity No studies reporting field validation have been located by the writer

Use of the Elwell-Fowlkes Bookkeeping Test in Selection The nature of its content, and its reliability, suggest that this test might be effective as a means of checking the mastery of bookkeeping fundamentals in inexperienced employment applicants Local norms are desirable, in view of variations in requirements and opportunity for learning on the job

Interview Aids and Trade Questions (827)

During both World War I and World War II extensive use was made of trade tests in the rapid classification and assignment of military personnel The first trade tests were described in detail by Chapman (154), those developed by the United States Employment Service have yet to be de-

scribed in detail Between the two World Wars the technique of trade testing was further developed at the Cincinnati Employment Center (827), where the trade questions were revised and brought up to date Subsequently the United States Employment Service recognized the value to this approach, and developed trade questions for use in its work, as described in Stead and Shartle (750 Ch 3 and pp 156–162) Because of its general availability, Thompson's Cincinnati work (827) is described here in order to illustrate the technique, it should be stressed, however, that occupational changes which have taken place since the mid-thirties make local and up-to-date revisions such as those of the USES necessary before his trade questions are put to practical use

Applicability, Content, Administration, Scoring, and Norms These trade questions were designed for use in employment offices and standardized on experienced craftsmen, but they may also be used with high school students who have had some trade training and are seeking their first employment The book consists of questions concerning the tools, materials, and methods of 131 trades ranging from Ammonia Pipefitter and Armature Winder to Wood Finisher and Woodmill Worker Each test contains from 15 to 25 questions, such as "What kind of weld has a boiler tube?" (for Boilermaker), the correct answer to which is "lap or nobble " The examiner reads the questions aloud, and they are answered orally The examiner notes the answers This procedure has the advantage of appealing to manual workers more than would a paper and pencil test The number of right answers is converted into a decile rating and a proficiency rating ranging from novice to expert Norms are occasionally based on small groups, the work having been published while in process of completion

Standardization and Initial Validation Because of the tendency to rely on internal validity in achievement and proficiency tests, and because of early publication of the book, statistical evidence of validity is lacking However, the fact that questions were developed with the aid of specialists in each field, and their ability to differentiate novices from journeymen and experts, constitute evidence of a sort

Reliability Data are not presented on the reliability of these trade questions Stead and Shartle (750) reported reliabilities of 79 to 93 for the USES tests

Validity No later studies casting light on the validity of these questions in selecting workers have come to this writer's attention

Use of the Interview Aids and Trade Questions in Selection Experi-

ence has repeatedly shown that a few well selected questions concerning the tools, material, and methods of his claimed trade are likely to weed out the ill-trained or inexperienced worker who wants to bluff his way into a desirable job (a very real problem in military classification) and command the respect of the expert who knows his craft The use of trade questions in employment offices therefore seems amply justified, even though controlled experiments and quantitative data are lacking

CHAPTER VIII

CLERICAL APTITUDE· PERCEPTUAL SPEED

Is There a Special Clerical Aptitude?

A COMMONLY used classification of clerical jobs (93) describes three phases of clerical work doing the work, checking it, and supervising it Job analysis has shown that these are levels as well as types and that each of these levels of clerical work requires the making of more decisions than the level immediately below it But planning and decision-making imply intelligence, aptitude for abstract thinking, a requirement by no means confined to clerical work This being the case, one might well ask whether there is actually such a thing as clerical aptitude Material discussed in the chapter on intelligence shows that general intelligence is indeed a factor in success in clerical work the minimum desirable I Q being 95 or 100, and the minimum requirements rising with the level of responsibility When promotability is a factor to be considered in the counseling or selection of potential clerical workers, intelligence should be heavily weighted, when, on the other hand, success in a routine clerical job is in question, intelligence exceeding the minimum requirement is all that is needed, other factors then being the decisive ones What these other factors are will be seen below

Job analysis suggests other aptitudes which should be important in clerical work In routine clerical work, at least, one would expect speed and accuracy in checking numerical and verbal symbols to be a characteristic of the successful worker Bookkeeping, typing, filing, and other record keeping jobs involve constant checking or copying of words and numbers, calling for perceptual speed and accuracy on the part of the employee It will be seen below, in the discussion of the Minnesota Clerical Test, that this hypothesis is borne out by research, it will also be seen that speed in perceiving numerical and verbal similarities is so much more important in clerical than in other occupations that there is some justification for referring to this ability as clerical aptitude

Another aptitude which job analysis suggests should contribute to success in clerical activities is motor skill or manual dexterity Standard works on aptitude testing such as that by Bingham (94 152) list motor skill as one of the aptitudes required in clerical work, with the obvious justification that such work involves frequent and rapid manipulation of papers, cards, pencils, typewriters, and other office tools and machines As will be seen in the next chapter, which deals with manual dexterities, the only evidence from aptitude tests to support this claim lies in the superior scores made by clerical workers on fine manual dexterity tests No studies of the relationship of such tests to clerical success are known, and gross manual dexterity has actually been demonstrated to be unrelated to clerical success It seems that other aptitudes such as intelligence and perceptual speed are so much more important that anyone who has average or better manual dexterity has enough motor skill for success To put it in other terms, the critical score for manual dexterity in clerical work is so low that almost everyone of average intelligence surpasses it

Finally, analysis of the work of office clerks has suggested that proficiency in language and in arithmetic is essential to success These are of course not aptitudes in the strict sense of the term, but only in the sense that such proficiency may be prognostic of success on the job However, it has been seen that the validity of clerical proficiency tests has not actually been demonstrated against external criteria, legitimate though the assumption may appear

The answer to our initial question is, then, that two or more aptitudes contribute to success in clerical work, and that one of these appears to be peculiarly important, partially justifying referring to it as clerical aptitude Although perceptual speed as measured by other techniques is important in other occupations (336), it has been shown that there are two perceptual factors, one involved primarily in the perception of space relations, the other primarily in clerical (numerical and verbal) tasks (735 152) The latter's importance in clerical work is such as to warrant its treatment as clerical aptitude The balance of this chapter will therefore be devoted to a survey of perceptual speed as clerical aptitude

Typical Tests

Tests measuring perceptual speed by means of numerical or verbal symbols have long been a standard part of the armamentarium of the psychologist, a number of them having been included in the grandfather of measurement texts, Whipple's *Manual* (919) It was not until the days

of more refined statistical procedures and validation against occupational success, however, that the peculiar value of these tests in vocational guidance and personnel selection became obvious The idea was tried out and validated as the *Minnesota Vocational Test for Clerical Workers* at the Minnesota Employment Stabilization Research Institute by Paterson and Andrew (see below) The Psychological Corporation's *General Clerical Test* and the *O'Rourke Clerical Test* incorporate items of the same type, together with others which measure numerical and verbal abilities more complex than mere perceptual speed The Minnesota test is the only clerical aptitude test which has been subjected to widespread and careful study and validation It is therefore the only instrument in this category to be discussed in detail

The Minnesota Clerical Test (Psychological Corporation, 1933, 1946)

This test was the one test construction project carried out by the Minnesota Employment Stabilization Research Institute, which found that its needs for tests of intelligence, manual dexterity mechanical aptitude, spatial visualization, and personality were fairly well met by the then available instruments It was so easy to administer and score and so thoroughly studied that it immediately became one of the most widely used aptitude tests It was originally called the Minnesota Vocational Test for Clerical Workers

Applicability, Effects of Age, Training, and Experience The Minnesota Clerical Test was designed and standardized for adult use, the adult group including girls of 17 and above and boys aged 19 and above It was then assumed that the test would be equally applicable to boys and girls of high school age, but data for age and grade norms were subsequently compiled (678) These show an increase in scores with age and grade, the median Number-Checking scores for 14, 15, 16, 17, and 18 year-old boys being 89, 94, 100, 104, and 102 As Schneidler points out, the sample is not perfect for age norms, as it includes only those who happened to be in grades 8 through 12 the duller 14 year olds were therefore not included, and the brighter 18 year-olds had already graduated from school However, the age and grade norms resemble each other enough to give one some confidence in both sets

Unfortunately, Schneidler's analysis is not sufficiently refined to answer the important question concerning the applicability of the test, to wit, that concerning the influence of age on scores Her data reveal an increase in the mean scores of increasingly higher age groups, but they do not

indicate whether this increase is due to the selection which normally operates in high schools to eliminate the less intelligent as they get older, to the maturation of clerical aptitude with age, or to the effects of training and experience in high school which involve practice in speed and accuracy of perception She does provide intelligence test data for her sample, but these are in terms of scores which are insufficiently described to permit interpretation If they are intelligence quotients, then there was no selection on the basis of intelligence, for the scores remained relatively constant throughout the four years of high school This would indicate that the increase in clerical scores with age is due either to maturation or to experience While it would be surprising to find so simple a skill maturing as late as the last two years of high school, it would be still more surprising, in view of data to be presented below, to find that experiences as dissimilar to that of the test as high school work affect the test scores There are clearly some important problems for further investigation here before this simple-appearing test is really understood In the meantime it must be used with caution at the adolescent level

An attempt was made by Klugman (435) to ascertain the effect of a year of *schooling* on the test His subjects were a group of 207 commercial high school girls, who showed significant gains in scores on both parts of the Minnesota Clerical Test after a year of high school commercial education As the 30 oldest did not differ significantly from the 30 youngest, Klugman concluded that the increase was due to training rather than to maturation To this writer the conclusion does not seem warranted, in view of Schneidler's prior finding that scores increased with age in all types of high school students It is regrettable that Klugman used no control groups

The problem of the *effect of experience,* as distinct from maturation, is not confined to the use of the test with adolescents Andrew (21) investigated it in the original studies with the test, administering it to 155 clerically experienced women aged 17 to 29 and correlating scores with amount of experience The correlations for the Numbers and Names Tests were 30 and 31, respectively This might be taken as indicating that clerical experience has some effect on Minnesota Clerical Test scores, were it not for one problem of sampling the less experienced group could normally be expected to include some relatively unselected workers of low aptitude who are normally weeded out during the first year or so of experience and who shift to light factory, sales, or other non-clerical employment If this group could have been sifted out, in

retrospective analysis, it might have left a group in which "true" clerical aptitude was equally distributed and in which the correlation between Minnesota Clerical Test scores and length of experience was zero In another study (22) Andrew administered the test to 28 clerically inexperienced adults before they embarked upon a five-month training program in clerical work, and readministered it again at the end of the training period The difference between pre-training and post-training scores was not significant, leading to the conclusion that training in clerical work had no effect on the scores of the Minnesota Vocational Test for Clerical Workers

A further study of the effects of experience was made by Hay and Blakemore (360) in a large bank They tested 229 inexperienced and 241 experienced women applicants for clerical employment The experienced group averaged 7 points higher on the Names, and 7 5 points higher on the Numbers Test, the equivalent of less than 25 sigma or 7 percentile points at the mean These differences are statistically significant, but in practice they are not likely to prove vital, especially if the reasoning applied to Andrew's first study is valid and applicable here Indeed, it is highly likely that Hay's inexperienced applicants included some women of little true clerical aptitude who would in due course be weeded out and who would not subsequently be in the market as experienced applicants for clerical employment If this is so, then it would be all the more legitimate to consider the small but statistically significant difference reported by Hay and Blakemore as psychologically and practically insignificant The authors found negligible correlations between scores and length of experience in clerical work, further supporting this conclusion

In summary, it seems necessary to conclude that it has not been demonstrated that training or experience affect scores on the Minnesota Clerical Test The preponderance of evidence from several ambiguous studies, together with the clear-cut findings of one study of the effect of training on scores, indicates that the test is relatively independent of training and experience in clerical work

Sex differences have been found to be significant (22,681) This means that although the test is usable with both sexes, separate norms are needed Women tend to be superior to men in general, although in the same job men and women are found to be equal in clerical aptitude, indicating the effects of selection *Age*, however, has no effect on scores according to evidence compiled with adult groups by the same authors

Content The Minnesota Clerical Test consists of two parts, the Numbers Test and the Names Test The former is made up of a series of pairs of numbers, in some of which the members are identical and in some different, as in the following samples 7639——7693, 6291——6291 The examinee must mark the pairs in which the two members are identical The Names Test embodies the same principle, with only minor differences between the members of non-identical pairs, e g Smith and Co ——Smyth and Co The task is obviously simple and routine in nature although exacting when speed and accuracy are required

Administration and Scoring The test is designed for group administration and requires fifteen minutes working time Examiners need to make sure that subjects are working on the proper part of the test, and that they draw a line, as directed, under the last pair at which they looked before the direction to stop was given Scoring is by means of a stencil, and involves a correction for wrong answers Scores are thus a combination of speed and accuracy, and have been criticized as such by Candee and Blum (134), who developed a scoring system which yields separate scores for speed and accuracy Their contention is that accuracy is more important than speed in clerical work, a slow accurate worker being preferable to a fast inaccurate worker Such a scoring method might be desirable when the criteria permit evaluation of the relative importance of each factor, but in most situations they are not so refined It seems probable that the combined score provided by the test authors is generally to be preferred for occupational use, giving as it does some weight to each factor The great majority maintain a fair degree of accuracy, knowing that it counts together with speed, and the important individual differences revealed in the test are differences in speed (171) If an examinee lowers his accuracy level in order to increase his speed, the wrong-penalty minimizes the gain

Norms The manual provides norms for gainfully employed adults, clerical workers in general, and various specific clerical occupations such as shipping clerks, routine clerical workers, bank tellers, and accountants and bookkeepers The general adult group is the standard sample used in the Minnesota Employment Stabilization Research Institute, a cross-section of 500 gainfully employed persons in the Twin Cities, so selected as to be representative of the urban national occupational distribution The norms for the specific clerical occupations are unfortunately not as satisfactory, consisting as they do of small groups of relatively undescribed workers in each category The accountant and

bookkeeper group, for example, includes 29 men, there is no indication as to how many were accountants and how many bookkeepers, a factor which has a bearing on their probable intelligence level, nor is evidence presented to show how representative they are of accountants and bookkeepers in general The 181 women stenographers and typists illustrate the same problem how many are secretaries, how many are stenographers, and how many are typists? The norms do not tell one Despite these defects, the original norms seem to have some validity, for they are not excessively out-of-line with those for the equally small groups studied by the United States Employment Service (Figures 4 and 5)

Fortunately research has not ceased with the compilation of the original sets of norms just referred to (and it should be stated parenthetically that at the time of publication, a mere fifteen years ago, these norms were unusually comprehensive) Norms and critical scores have been made available for more than 700 men and 1400 women bank employees by Hay and Blakemore (359), and adolescent norms have been published by Schneidler (678) and discussed elsewhere in this chapter Both are included in the revised manual The median scores for clerical workers in the Philadelphia bank studied by Hay and Blakemore were about ten raw-score points below the mean reported by the Minnesota studies for routine clerical workers, and equivalent to the 85th percentile (men) and 70th (women) when compared to the general adult sample of the Minnesota project Hay found a critical score of 130 (Numbers) useful in selecting machine bookkeepers this is about the median for routine clerical workers according to the Minnesota norms, and 19 raw-score points below the Number-Checking median for Minnesota office-machine operators Since it does not seem likely that Philadelphians are inferior in perceptual speed to Minnesotans, and since the former sample was collected over a period of years which included both depression and prosperity, whereas the latter was taken at the depth of the depression when inferior clerical employees had been released (22), it seems likely that Hay's norms are more representative This is confirmed by a USES study cited below It is noteworthy, however, that the critical score which Hay established for his concern was almost identical with the median for the employed routine clerical group in the Twin Cities and for one of the USES samples The median and critical score on the Otis S A Test for Hay's clerical workers being an I Q of 100, and the first quartile an I Q of 95, it seems legitimate to treat his sample as about the same as a routine clerical group The writer is in-

clined to wonder about the wisdom of critical scores which, like Hay's, are the same as the median of a successfully employed group, a Number Test score of 112, about one sigma below the mean, would normally be more practical

Although not presented primarily as norms, the USES data published by Stead and Shartle (750) and reproduced in Figures 4 and 5 provided a valuable source of norms for the Minnesota Clerical Test In these figures the means and standard deviations of the raw scores made by various types of clerical and semiskilled workers on both Numbers and Names Tests are graphically presented Although the numbers are small the data agree reasonably well with those of the Minnesota studies and with Hay's, guidance in terms of the limits suggested by the first sigma points of the USES data, and, lacking local norms, selection on the same basis, will probably not be far wrong

The Minnesota Clerical Test has now been sufficiently widely used for more adequate norms to be available for specific clerical occupations

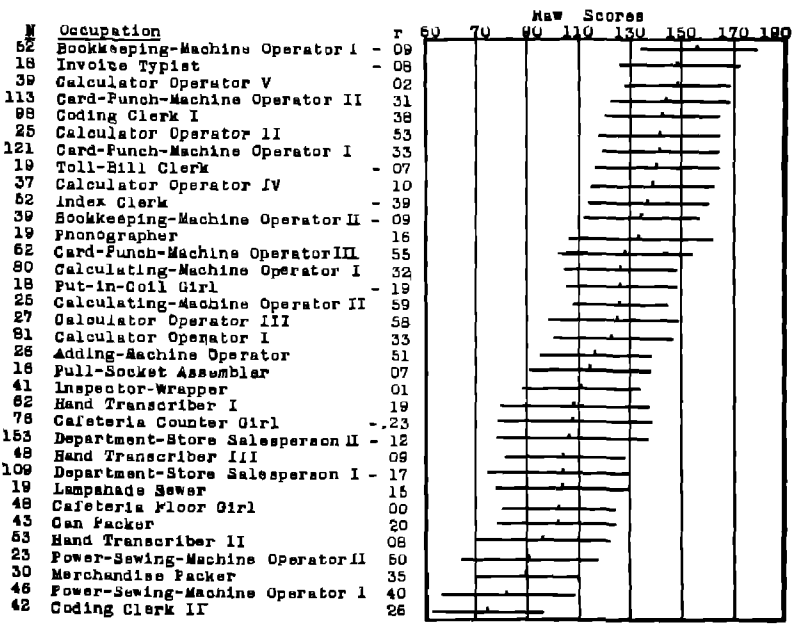


FIGURE 4

OCCUPATIONAL DIFFERENCES ON THE MINNESOTA CLERICAL NUMBERS TEST
Means and Standard Deviations after Stead and Shartle (750)

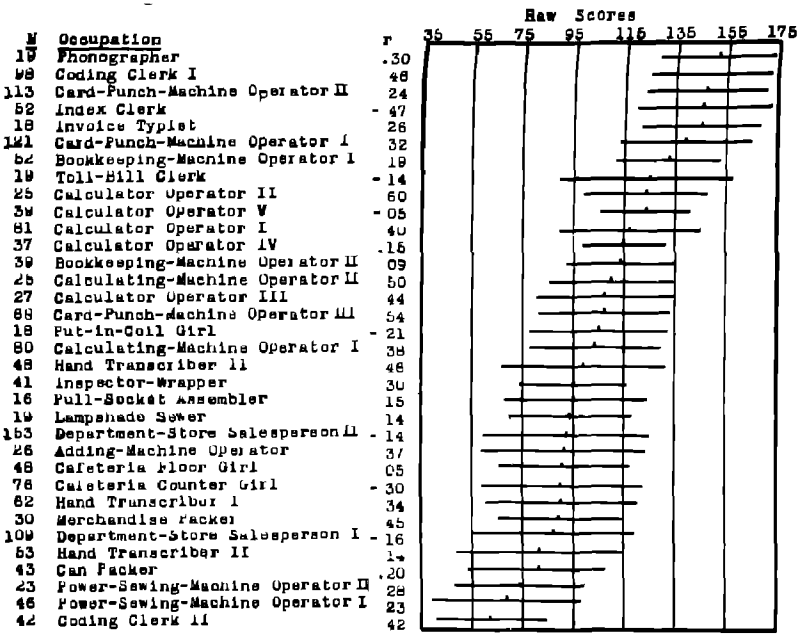


FIGURE 5

OCCUPATIONAL DIFFERENCES ON THE MINNESOTA CLERICAL NUMBERS TEST
Means and Standard Deviations after Stead and Shartle (750)

The latest (1946) manual includes norms from all the above groups except the USES subjects Norms for students in a graduate school of business have been published by Strong (781) With the advances that have been made in selecting and describing samples ever since the test first appeared, it is to be hoped that future editions of the manual will describe even more adequately the groups used in norming the test

One other problem remains to be discussed in connection with the norms, stemming from the age differences which have already been considered There is a very real question as to which *norms* to use when counseling *high school students*, a problem which Barnette (44) has also encountered with business college students This may best be illustrated by a specific example An 18 year-old high school senior, let us say, is considering taking training to be an accountant, has taken the Number-Checking Test, and made a raw score of 106 This puts him at the 50th percentile for his grade, the 58th for his age, and the 74th when compared

to employed adults So far, then, the picture is one of average or superior clerical aptitude, although one might suspect that the superiority of an average high school senior when compared to employed adults is the result of the selectivity of high schools However, when compared to accountants and bookkeepers, the group with which he is to compete, his percentile rank drops to the first The counselor must ask himself whether this is his true and ultimate standing when compared with accountants and bookkeepers, in which case he should certainly be encouraged to consider other possibilities, or whether the poor standing is the result of immaturity and therefore subject to modification by age and maturation If the latter is the case, then he and all of his fellow seniors will improve in score, making them even more superior to adults-in-general, although it does not seem likely that they should actually exceed more than 75 or 80 percent of the employed adult population in clerical aptitudes In view of this last consideration, it seems wise to assume that there will not be much change in the raw scores of high school seniors after graduation (assumption supported also by the lack of relationship between age and scores among persons aged 17 to 19 and above previously mentioned) The adult occupational norms should therefore be used cautiously even for high school juniors and seniors, rather than the age or grade norms made available by Schneidler

When in due course more light is thrown on the role of maturation it may be shown to be necessary, and it may become possible, to provide conversion tables which will show the probable adult score of an adolescent who has made a given raw score, by converting the adolescent raw score to the adult equivalent, and this to the specific occupational percentile, one will then be able to make a fair evaluation of an adolescent's prospects of successful competition in a specific clerical occupation

Standardization and Initial Validation The several times revised manual for the Minnesota Clerical Test has been more complete than most in the presentation of data concerning the standardization and initial validation of the test, and has gone somewhat beyond that in summarizing subsequent findings—a pattern now fortunately being increasingly followed by the more responsible publishers and authors of tests The data which follow concerning the standardization of the test are therefore also found in the manual

The *correlation between Number-Checking and Name-Checking* was found by Andrew (21) to be 66, indicating that the tests have a great

deal in common but that, since their intercorrelation is lower than their individual retest reliabilities of .76 and .83 (187), at least one of them is measuring something not so well measured by the other.

This was shown by other correlational data to be *intelligence,* which plays a more important part in Name-Checking than it does in Number-Checking. In homogeneous groups the correlation between the former and the Pressey Senior Classification and Verification Tests (of intelligence) was found to be .37, whereas the same figure for the latter was .12. In heterogeneous groups these correlations rose to .65 and .47. These data bring out another fact important to an understanding of the nature of clerical aptitude. In a group of persons of the same general level of intelligence, such as one normally finds in a class in a large high school and in a business office, clerical perception is an aptitude which is unrelated to intelligence, on the other hand, in a group of persons with a wide spread of intelligence, such as one finds in a class in a smaller high school where sectioning has not been possible or in a group of unsorted applicants for clerical employment, those who are more intelligent tend to have more clerical aptitude than those who are less intelligent. The relationship is far from perfect, but it is real.

As the test involves reading words and numbers Andrew correlated it also with tests of *reading speed, spelling,* and *arithmetic.* Using homogeneous groups, the correlations between reading on the one hand and Numbers and Names on the other were respectively .09 and .45, the correlation between Names and spelling was .65, that between Numbers and arithmetic was .51. Holding intelligence constant, since it plays a part in reading and in the Names Test, the correlations between reading and the Numbers and Names Tests changed to .18 and .30. Since reading and arithmetic are proficiencies and perception an aptitude, one would be inclined to assume that clerical aptitude explains the reading and arithmetic scores, were it not for the fact that skill in reading affects the speed of perception of symbols such as those used in the Minnesota Clerical Test. Leaving the riddle of the hen and the egg unsolved, it is still possible to conclude that, in homogeneous groups, the relationship between reading skill and clerical aptitude is relatively low. In the case of arithmetic the riddle is more solvable, for the Numbers Test requires no computation and is therefore not affected by proficiency in arithmetic; the relationship reported must therefore be causal from Number-Checking to computation rather than vice-versa. This may perhaps justify the

conclusion, by analogy, that speed of reading is affected by perceptual speed as measured by Name-Checking

Andrew attempted to ascertain the relationship between *training and experience,* on the one hand, and Minnesota Clerical scores on the other As these studies have been discussed earlier in this chapter they will not be dealt with here

The relationship between clerical aptitude and *success in clerical training* was ascertained (22) More than 100 commercial high school students were rated for prospects of success in training by their teachers, the ratings correlating 58 with total Minnesota Clerical scores and 43 with intelligence test scores The correlations with college accounting grades were found to be 47 for Numbers and 49 for Names (22) These results seem extraordinarily good unfortunately, it will be seen that subsequent field validation has not tended to confirm them

The validity of the test for selecting *clerical employees* was ascertained (22) by correlating supervisors' ratings with Minnesota Clerical Test scores The groups involved ranged in size from 22 to 97 workers, the reliability of the ratings was not checked Even with this presumably imperfect criterion, the test validities ranged from 28 to 42. Subsequent studies of the same type, discussed below, have yielded similar results

Employed and unemployed clerical workers were compared (22) in order to ascertain whether or not there were measurable differences in clerical aptitude between such groups The critical ratios were 3 32 for Numbers and 4 49 for Names, showing that the employed clerical workers were significantly superior to the unemployed clerical workers on these tests Further analysis showed that the early unemployed were inferior to those who had been released later in the depression, as well as to the still employed, but that the late unemployed were not inferior to the still employed As it seems logical that the first to be released would be those whose services were least valued by employers, and the last those whose services were difficult to dispense with, this would seem to be a validation of the Minnesota Clerical Test against employers' ratings of essentiality an efficiency rating made much more carefully than the average rating

A final type of preliminary validation of the test carried out by the Minnesota Employment Stabilization Research Institute was the ascertaining of the ability of the Minnesota Clerical Test to *differentiate clerical from non-clerical workers* (22) This involves the hypothesis

that the trait measured is truly an aptitude, for the acceptance of which evidence has been adduced, and the further hypothesis that the aptitude is not so widely or generally distributed that people-in-general possess it in a degree equal to that characterizing those in the occupation, hypothesis which is automatically checked in this type of validation

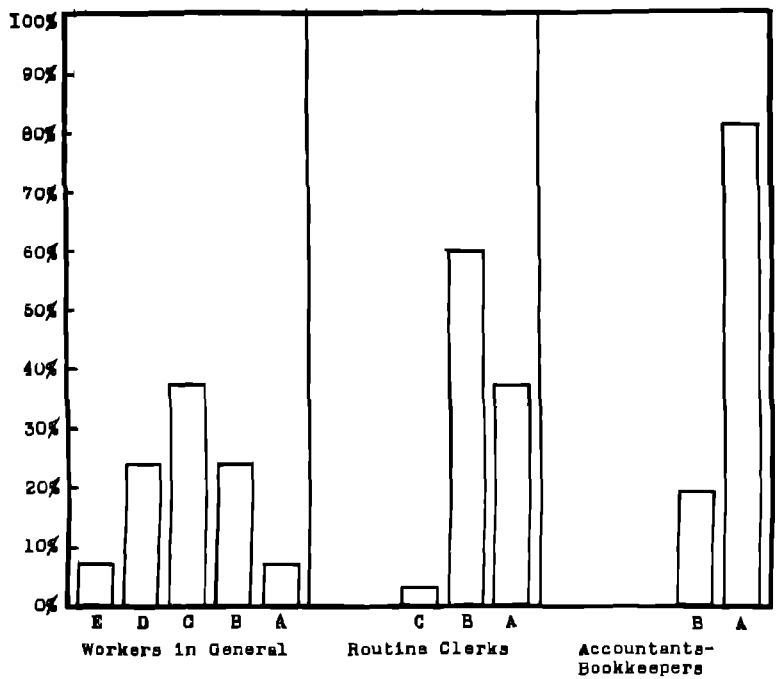


FIGURE 6

OCCUPATIONAL DIFFERENCES ON THE MINNESOTA CLERICAL (NUMBERS) TEST
Showing the percentage of each type of worker making a given letter grade After Andrew and Paterson (22)

Figure 6 reproduces data from the MESRI studies (22) which graphically portray the ability of the Minnesota Clerical Test to differentiate between workers-in-general and workers employed in various clerical occupations The distribution of scores for men-in-general is normal, whereas the higher one goes in the scale of clerical occupations the more skewed the distribution becomes Approximately 7 percent of the worker-in-general group received letter ratings of E on the Numbers Test, while no routine clerical workers received a grade of E, in fact, none of the

latter group received ratings of D, and only 3 percent received a grade of C Accountants and bookkeepers, on the other hand, in no cases received a rating as low as C, and more than 80 percent of them rated A on the Numbers Test, as contrasted with about 37 percent of the routine clerks and 7 percent of the workers-in-general

Although the differentiation between clerical and non-clerical workers shown above is striking, it should not be taken as indicating that no non-clerical occupational groups excel in what has for convenience been labelled clerical aptitude As the Minnesota norms bring out, miscellaneous minor executives, life insurance salesmen, retail salesmen, and draftsmen are all above the 80th percentile on the Numbers Test This is perhaps only to be expected, in view of the fact that in all of these occupations there is a great deal of record keeping or work in which minute details must be accurately and quickly checked Even policemen rate at the 66th percentile But these scores seem less impressive when it is noted that the only male clerical group whose median is below the 91st percentile when compared to the general population is the shipping and stock clerk category at the 77th percentile

Reliability The corrected split-half reliabilities were found to be 85 for the Numbers Test and 89 for the Names Test (manual), while the retest reliabilities were somewhat lower, 76 and 83 respectively (187) Hay (358) found retest reliabilities of 61, 69, and 56 for the Numbers Test, and of 75, 62, and 81 for the Names Test after intervals of as many as 54 months

Validity Because of its rapid and widespread adoption a number of validation studies have been carried out and published by workers in the field These studies have included the usual variety of correlations with other tests and with educational and vocational criteria

The relationship between Minnesota Clerical Test scores and *intelligence* was checked by Copeland (171) and Super (792) In the former study, correlations with the Otis S A Test were found to be 34 for Numbers and 51 for Names, in the latter study the A C E Psychological Examination was used, and the correlations were 26 and 62 respectively The range of intelligence and clerical aptitude was probably greater in the latter group, which consisted of high school juniors and seniors, than in the former, which was made up of unemployed clerical workers This would explain the closer relationship between intelligence and the Names Test in Super's study, but not the slightly lower relationship with the Numbers Test, which may be due to chance Both of these

relationships are in-between those reported by Anderson and Paterson for more strictly homogeneous and heterogeneous groups Tredick (869) correlated the Numbers and Names Tests with Thurstone's Primary Abilities Tests The former had low r's with verbal, memory, induction, and reasoning tests (06 to 28), the latter with only the memory test (24) Other r's were above 36 Both tests are heavily loaded with perceptual and numerical factors, while the Numbers Test is weak in the verbal and reasoning factors

The relationship between the clerical test and the Co-operative Survey Test in Mathematics was computed in an unpublished study of high school juniors and seniors with negligible relationships resulting (— 07 and — 10) This does not seem to agree with Andrew's finding of 51 for Numbers and arithmetic, but may be due to the more advanced mathematical content of the Co-operative test, which requires reasoning more than routine computation

The relative validity of the Minnesota Clerical Test and the General Clerical Battery of the United States Employment Service was ascertained by Ghiselli (285), who administered both to a group of 562 workers His analysis showed that the latter added nothing to the former, which was adequate for counseling use

Teachers' ratings of written work were used as a criterion by Swem (809) His subjects were 35 boys and 39 girls enrolled in high school courses For the former the correlations with Numbers and Names Tests were 30 and 19 respectively, for the latter they were 05 and 31 Only the correlations for the Names Test were statistically reliable These findings contrast unfavorably with those reported in the original studies

The relationship between Minnesota Clerical scores and *grades in typing and shorthand* was analyzed by Barrett (46), working with groups of 96 and 75 college students Unfortunately her analysis was not made in terms of correlations or similar statistics, but inspection of her data shows a tendency for those who made higher scores on the Minnesota test to make higher grades in both typing and shorthand Tredick (869) found correlations of 08 31, and 27 between grades in Art, Chemistry, and English Composition on the one hand and Numbers on the other, the figures for Names were 07, 26, 07 Correlations with average grades were 36 and 23 The subjects were 113 freshmen women in Home Economics

An *examination in machine calculation* was used as a criterion with

51 women students of office practice by Gottsdanker (303), whose battery of tests included a slightly modified version of the Numbers Test The correlation between his Number-Comparison Test and the criterion was 29, when combined with the Tapping Test of the MacQuarrie Mechanical Ability Test, a Number-Dot Location Test (a "paper keyboard"), and an Arithmetic Computation Test, the multiple correlation coefficient was 57

Output on the job served as criterion in a study of 39 bookkeeping machine operators by Hay (358) Speed of posting was used as an index of output for, as Hay points out, operators are not permitted to remain at work if they make errors and so learn to work at an accurate speed, making speed the best index of success The reliability of the production trials was checked, and was found to vary around 90 for any one trial period, when inter-trial reliability was checked, it was somewhat lower, but in no cases lower than 72 With this carefully studied criterion the correlations for Numbers and Names Tests were 51 and 47, respectively When these two tests were combined with the Otis S A Test a multiple correlation of 65 was obtained Hay has used this battery in a large bank for a number of years with cut-off scores of 130 for the Minnesota tests (359)

Supervisors' ratings of the efficiency of clerical workers were used as criteria in another study (193), in which the validity of Numbers and Names Tests was found to be 27 and 29 When promotability was estimated by job level attained after five or more years of service and correlated with the same tests, the coefficients were 07 and 34 The Thurstone Examination in Clerical Work (a proficiency test), and a test of the same type by O'Rourke had validities ranging from 40 (efficiency) to 77 (promotability) This is not the reflection on the Minnesota test that it might seem at first glance, because the former instruments are tests of mixed functions, comparable to a battery, whereas the Minnesota is a purer test of two factors only, perceptual speed and, to a lesser extent, intelligence It is to be expected that tests of clerical tasks would correlate more highly with efficiency ratings than a test of perceptual speed, and that tests as heavily loaded with intelligence factors as the Thurstone and O'Rourke would correlate more highly with promotability When selecting new workers, however, there are important advantages in using a battery of purer tests, one of intelligence, one of perceptual speed, and one of arithmetic or language usage, depending

upon the type of clerical work In Hay's work (358), for example, the first two proved sufficient, because in that case neither arithmetic nor language was of special importance

In the USES study (750), data from which are reproduced on pages 169 and 170, a battery of tests including the Minnesota Clerical Test was administered to various groups of clerical workers For two samples totaling 234 card punch machine operators (sex unspecified but presumably female) the criterion was the average number of cards punched per hour, with each incorrectly punched card counted as one error, combined into an "errorless production" score The reliabilities of the two components, cards punched per hour and number of errors, were about 95 for the former and 90 for the latter Coding clerks (N = 96), bookkeeping-machine operators (N = 52) and hand-transcribers (N = 62) were studied with a similar battery and criterion, for calculating-machine operators (N = 80) and adding-machine operators (N = 26) the criterion was a worksample For card-punch machine operators the validities were 31 and 33 for Numbers, and 24 and 32 for Names, these were among the most valid tests in the battery, only a letter-digit substitution test being as good and the MacQuarrie subtests having no consistent validity The validities for coding clerks were 38 and 16 for Numbers and Names, validities equaled by a number writing test and exceeded by a personal data test In the case of the bookkeeping machine operators the validities were — 09 and 19 although, as will be seen later, this group tended to make high scores on the tests and Hay (758) found validities of 51 and 47 perhaps the difference lies in the criteria, the USES having used an error criterion while Hay used a speed criterion which he considered more valid That Hay's criterion was superior is suggested by the relatively low validity of the other tests in the USES battery, none of which exceeded 28 For the hand transcribers the coefficients were 20 and 34, again among the best of the battery, sentence-completion, vocabulary, and number-writing tests being in the same range Validities for calculating machine operators were 34 and 38 for Numbers and Names, for adding-machine operators, 51 and 37 For the former group MacQuarrie Tracing and Location, and a number finding test were also valid, for the latter, all of the MacQuarrie subtests except Tapping and Dotting had some validity, as did vocabulary number-finding, and an arithmetic test

Data for other clerical groups, including some comparable to those just discussed, and for a number of semiskilled jobs in which it was

assumed clerical perception would be important, are reproduced in Figures 4 and 5, pp 169 and 170 Worthy of note are the substantial negative correlations between Numbers (— 39) and Names (— 47) and the ratio of errors to the production of index clerks, whose average scores are more than one sigma above the women's mean this suggests that in this occupation a high level of perceptual ability is desirable, but that those who are too much above the critical minimum are likely to be the poorer workers whether or not this is because their rate of work is too fast for the precision requirements of the job is not shown by the data Also noteworthy, in view of the Blum and Candee study cited below are the correlations of 015 and 30 between Numbers and Names on the one hand and ratings of inspector-wrappers on the other, and those of 35 and 45 between the same tests and production records (ratio of time required to standard time per unit) of merchandise packers Another non clerical job for which the test had some validity was power-sewing machine operator (40 to 50, and 23 to 28)

Blum and Candee (106) tested 317 seasonal and 55 permanent *packers and wrappers* in a department store In the permanent group the Numbers Test had a correlation of 57 with packers' production, and the Names Test one of 65 with wrappers' production In the seasonal group only manual dexterity was important The authors' conclusion that the initial job adjustment of packers is somewhat affected by speed of gross arm and-hand movement, while long-term superiority is more dependent upon clerical speed and accuracy, seems legitimate But the differential results for Numbers and Names, packers and wrappers, need further investigation before the matter is closed the USES study showed that both were valid for packers

In the study of pharmaceutical *inspector-packers* Ghiselli (286) worked with 26 young women who were rated by their forewoman and supervisor The correlation between the two sets of overall ratings was 72, which was considered adequate reliability and justification for combining the two to serve as criterion The correlations with Minnesota Numbers and Names Tests were respectively 29 and 26

Apparently packing work of both gross (department store) and fine (pharmaceutical) types requires speed and accuracy of perception such as is measured by the Minnesota Clerical Test Just why the gross type should require it more consistently than the fine is difficult to see It will probably not become clear until other studies of these and other packing jobs are made with the same tests, in combination with detailed

job analyses It would be illuminating, for example, to know whether it is speed in recognizing numbers and names, as in the Minnesota test and in clerical activities, which is important in packing and wrapping, or whether it is general perceptual speed and accuracy such as might be measured by other speed of discrimination tests If the former, the Minnesota test is perhaps truly a test of clerical aptitude, if the latter, it is more probably a perceptual test measuring something of value in a variety of occupations The data on power-sewing-machine operators suggest that it is the latter

Two new studies have checked the ability of the Minnesota Clerical Test to *differentiate between persons* in clerical and non-clerical occupations In one investigation Barnette (44) found that business college students were superior to general adults, but inferior to clerical workers, on both Numbers and Names Tests One would expect this of a student group, some of whom were likely to be weeded out before establishment in the occupation, unless they were preparing exclusively for the higher levels of clerical work

The other study is from the United States Employment Service studies in occupational analysis, previously discussed, and cited by Stead and Shartle (750 Ch 8 and pp 217–225) As the sex composition of the occupational groups is not specified, comparing them with general population norms involves the assumption that most of the clerical workers studied were women, this is probably a legitimate assumption When comparing one clerical group with another the procedure is made more justifiable by the fact that men and women in a given clerical job are found to be equal in clerical aptitude As Figures 4 and 5 (pp 169 and 170) show, there is a definite tendency for clerical workers to make higher scores on the Numbers Test than workers in the semiskilled occupations to whom the same test was administered The mean scores of almost all of the clerical jobs tested were above the mean of the MESRI standard sample of employed adults, hand transcribers and one sample of coding clerks being the only clerical workers whose average is lower than the adult average A cut-off score of 122 (about one sigma above the adult women's mean) would include all of the clerical workers above the mean of their group except those just mentioned and ten-key adding-machine operators, of the 12 non-clerical jobs included in the list, only the put-in-coil girls have a mean score as high as this If Hay's critical score for bookkeeping-machine operators (also his mean) of 130

were used, all of the above-average bookkeeping-machine operators and other comparable clerical workers would surpass the critical score

The data for the Names Test show similar trends, although Hay's cut-off score of 130 appears to be too high for this test 105 or 110 would be comparable to that used for the Numbers Test, although the latter is about the mean for adult women The differentiating power of the Minnesota Clerical Test revealed by these data is greater than it at first seems, because the non-clerical jobs in the occupational sample were included on the assumption that perceptual speed as measured by the Minnesota test would be important in them too, hypothesis proved valid for some by the reported validity coefficients It is noteworthy, however, that these non-clerical jobs in which clerical perceptual speed is important almost invariably rank lower in the amount typical of their workers than do the clerical jobs themselves

One of the objectives of vocational counseling and selection is the attainment of *satisfaction in his work* by the worker This being the case, one would expect to find studies of the relationship between clerical aptitude and job satisfaction No such studies have been located, however, the emphasis having so far been entirely on success

Use of the Minnesota Clerical Test in Counseling and Selection The preceding discussion has brought out the fact that the Minnesota Clerical Test has value for distinguishing those who have promise for clerical work from those who do not, and that the higher the score made by a person the higher, other things being equal, he may rise in the field of clerical work Even though persons in the highest level clerical jobs are characterized by more perceptual ability than those in lower-level clerical work, one is not justified in assuming that this is all that need characterize the aspirant to high-level clerical work We have seen also that while perceptual speed is more important in routine clerical work than is intelligence, intelligence is probably more important in promotion to the higher levels than is perceptual speed

When appraising clerical promise it is well, therefore, to use tests of both perceptual speed and intelligence If a battery can be used, it should include the Minnesota Test (Numbers and Names) and an intelligence test such as the Otis If time is at a premium, the Minnesota Numbers Test and the Otis will do If only one test can be used, and it must be brief, then the Minnesota Names Test, as a combination of perceptual speed and intelligence, may suffice In selection programs, if the selection

is to be made from a wide range of ability an intelligence test may suffice as a screening instrument, because of the correlation between the two aptitudes in heterogeneous groups But if the selection is to be made from groups with a limited spread of general intelligence, the Minnesota Numbers Test is preferable as a purer measure of the important variable Although the differences between experienced and inexperienced workers on the Minnesota Clerical Test were slight, and probably due to selective factors rather than to experience, it is worth noting (until more conclusive evidence is available), that at least one personnel worker (Hay) has thought it advisable to use separate norms in selecting experienced and inexperienced clerical workers

In counseling the principal problem which is raised by the research is that of age and occupational norms Although increase in scores with high school grade and age has been demonstrated, it is not clear to what extent this is due to maturation and to what extent to the elimination of the less able students as they reach the higher grades In view of the fact that there is no change in scores with age from ages 17 to 29, and since the age changes in mid-adolescence are open to some question, it seems wise to use adult norms even with high school juniors and seniors until more adequate evidence is available on the effects of maturation When the test is being used at the junior high school level for curricular guidance purposes grade norms are to be preferred, as maturation may play a significant part at that age and school work can provide an exploratory experience which supplements the test scores Obviously, students who take commercial courses in high school should have appropriate mental ability and more than average clerical aptitude Since directional guidance is all that is needed at that stage, the more specific decisions can be postponed until a later age when tests and experience yield more specific evidence

In using the adult norms, emphasis should be on the occupational rather than on the general norms It only confuses the issue to know that a man is at the 74th percentile in accounting (Number) ability compared to men-in-general, when in reality he exceeds only 1 percent of accountants in that type of ability, for it is against accountants rather than men-in-general that he must match his accounting aptitude However, these occupational norms must still be used with considerable caution, since they are based on small and relatively nondescript groups whose representativeness is unknown except for the rough correspondence of MESRI, USES, and Hay's norms Most guidance centers should be able to develop

norms of their own which are more adequate for local use than the original Minnesota occupational norms, but these should be *occupational* norms, not norms for all clients locally tested

In administering the Minnesota test for non-clerical purposes as in the counseling or selection of semiskilled workers, it is well to supplement the directions with a statement that the test is a measure of the speed with which details are noticed, and that this is an important characteristic in a number of assembly, inspection, and other jobs This helps to counteract the antipathy of some examinees to anything with a clerical label

Finally, a word concerning speed and accuracy We have seen that as a rule speed on this type of test is a good measure of accuracy But there are occasional exceptions, and one subject will make a given score by working rapidly with errors, whereas another will make the same score by working more slowly without errors For this reason the psychometrist or counselor should examine the responses to each test, and take the error score into account in making his interpretation While it may not help as much in judging prospects of success as the total corrected score, it will help considerably in understanding the person being evaluated or counseled

CHAPTER IX

MANUAL DEXTERITIES

Nature and Role

Singular or Plural? Personnel men, vocational counselors, and psychologists have long been in the habit of referring to manual dexterity as though it were a unitary ability If this were so, then it would be legitimate to conclude that a person who is adept at one manual activity has the aptitude to become equally adept at any other manual activity It would also be true that one good test of manual dexterity would be sufficient in a battery used to survey the assets of a student or employment applicant

The plural form has been used in the title of this chapter in order to stress the fact that the research of the past decade (735) has demonstrated the existence of at least two types of manual dexterity gross and fine Another way of describing them might be as manual dexterity and finger dexterity, or confusion might be avoided if the terms arm and-hand and wrist-and-finger were substituted for these Further study may in due course reveal that even this breakdown is inadequate, and that dexterity is in reality a continuum, gross at one extreme and fine at the other, at least in a logical sense The use of different anatomical parts in gross and fine manual activities may, however, justify treating arm-and-hand and wrist-and-finger dexterities as discrete aptitudes It will be seen below that, at least as measured by the tests now available, these two types of dexterity are relatively distinct and unrelated to each other Furthermore, a factor analysis study of 59 different aptitude tests conducted by the United States Employment Service (735) revealed two dexterity factors, one of which was common to the Placing and Turning Tests of the Minnesota Rate of Manipulation Test and to the Peg Board Apparatus of the USES (both of which require relatively gross movements), and the other important in tests requiring fine assembly work Most relevant of all are studies by Seashore (698) and Buxton (127), using laboratory tests, in which factors which appear to consist of manipulative, wrist-turning, arm-and-shoulder, ballistic (uncontrolled), steadiness,

and one unidentifiable motor skills were isolated The first three appear to be distinct, anatomically based, dexterities No tests have yet been devised which suggest intermediate degrees of fineness, although some have been investigated which require varying combinations of arm-and-hand and of finger dexterities

What is Manual Work? Another set of distinctions which needs to be made early in the discussion of manual dexterities is those between manual work and mechanical work, manual dexterities and mechanical aptitude White collar workers and professional people who have not had intimate contact with industry often confuse manual and mechanical work and skills, taking note only of the fact that both involve use of the hands Aware that some factory and shop work is skilled, some semiskilled, and some unskilled they assume that these distinctions in the degree of skill characterizing the work are distinctions in degree of manual skill Hence the unwarranted conclusion that the higher the level of skill in industrial employment, the greater the need for manual dexterity

As experienced industrial men and personnel psychologists have long known, nothing could be further from the facts The independence of measures of manual dexterity and of mechanical comprehension or spatial visualization will be brought out in subsequent parts of this chapter and in the two which follow, as will the different degrees to which manual (unskilled and semiskilled) workers on the one hand and mechanical (skilled) workers on the other hand tend to possess these aptitudes It should suffice to point out here that "manual" work is essentially semiskilled or unskilled, semiskilled work relies primarily on the manual skill of the worker in assembling objects, packing them, or in other ways manipulating them with fingers or arms and hands, and unskilled work depends primarily upon the strength of back and legs and body co-ordination rather than eye-hand co-ordination, skilled work, on the other hand, is more dependent on the understanding and planning of the worker than upon mere manual dexterity To put it in everyday industrial parlance, the skilled worker needs "know-how," the semiskilled worker skillful hands and fingers, and the unskilled worker a strong back

A unique contribution to the understanding of manual skill and the nature of semiskilled work was made by Cohen and Strauss (162) in a study of 21 experienced women employed in a highly repetitive operation The task consisted of folding an 18 × 18-inch gauge sheet to a size

approximately 4 × 4 inches, and required six foldings Motion pictures were taken of the operatives at work, and operation analysis was made. It was found that, in general, the more skilled operatives (so classified by a standard time-and-motion study technique) performed their work more simply This greater simplicity of technique was illustrated by several differences in methods Better operatives have fewer limiting grasps and releases, in that they grasp and release as a part of transport operations rather than as separate movements, their movements are more global, less discrete, than those of inferior operatives The more proficient operatives make the movements of their two hands overlap more than the less proficient workers, thus performing two operations at once instead of one after the other The Purdue Pegboard, described later in this chapter, is almost unique in testing this type of two-hand coordination Poorer operatives make more extra moves because of fumbles, faultily performed operations, and superfluous operations than do the better operatives, the latter therefore have a shorter work cycle than the former, and a higher rate of production

Superior skill manifested itself not only as greater speed of performing basic operations but, the above makes clear, as improvement in the series of basic operations performed The authors therefore asked, "Is method independent of skill?" Their answer is an affirmative for general method, but a tentative negative for the basic operations An illustration helps to make the point "One operator releases a part during a motion rather than after it has been made, but if the less skilled operator attempts to do so, the part may not be placed correctly and an adjustment may be necessary Therefore the first operator can perform without the occurrence of 'Release' as a limiting operation, but the second cannot" (162 152) It is the accumulation of such small differences which differentiates operatives Cohen and Strauss feel (without evidence) that the problem is primarily one of selection rather than of training, and suggest that dexterity tests are needed which can measure the ability to eliminate limiting motions or to merge them into more global movements Although no available dexterity tests yield scores of this type, Test IV of the Purdue Pegboard (see below) provides excellent opportunity to observe this type of skill, and other dexterity tests give some clues

Typical Tests

The best known test of arm-and-hand dexterity is the Minnesota Manual Dexterity Test, better known as the Minnesota or Ziegler's Rate of

Manipulation Test No other test of this type has been widely studied or used. Wrist-and-finger dexterity tests include O'Connor's Finger and Tweezer Dexterity Tests and the Purdue Pegboard, the latter also a measure of arm-and-hand dexterity Both are dealt with in this chapter. Other tests of this type are the Pennsylvania Bimanual Worksample (Educational Test Bureau) and the pegboard and plier-dexterity tests of the United States Employment Service These and others like them are not treated here, because they are newer and less well validated or not generally available

The Minnesota Rate of Manipulation Test (Educational Test Bureau, 1931)

The Minnesota Rate of Manipulation Test was originally developed by Ziegler as the Manual Dexterity Test, in connection with a study of the role of manual dexterity in performance on the Minnesota Spatial Relations Test For this reason it was not, unfortunately, included in the Minnesota Mechanical Abilities Project (588), although several other tests designed to measure dexterity were used, it was, however, available in time for inclusion in the research of the Employment Stabilization Research Institute (589) It has been published in two editions, one by the Mechanical Engineering Department of the University of Minnesota, the other by the Educational Test Bureau The latter version differs from the former in the arrangement of parts at the beginning and the end of the test, in the number of parts (60 vs 58), and in the colors used on the movable parts, as the university version was used in the extensive normative work of the MESRI only it should be used with the employed adult and special occupational-group norms gathered by the Institute This fact appears to have been disregarded by the publishers of the other version, who give norms for 500 unidentified adults which seem to be those of the Minnesota project The Educational Test Bureau version is more widely used despite this fact, probably because of a more finished manufacturing job which includes a tray to hold the formboard and parts, combined with more aggressive marketing methods Supplementary norms for this form of the test are available, in the literature, as will be seen below, but the manual has not been revised in the necessary detail

Applicability The Minnesota Rate of Manipulation Test was designed for use with and standardized on adults It has generally been assumed that it is applicable at any age level between 13 and 50 (94), dexterity being a characteristic which matures relatively early However, the old

Educational Test Bureau norms show that men and women are faster than boys and girls, and Tuckman (877) found even greater differences between adults and adolescents According to his data, for example, a raw score of 232 5 is equal to the 50th percentile for boys, but the 27th percentile for men The question is raised as to whether these differences are due to the selection of the samples (clients of a guidance center may come for different reasons, from different backgrounds, at different ages), to differences in the motivation of the two age-groups (the boys may consider manual tasks beneath them while the adults are more realistic in their vocational objectives), or to the role of maturation (manual dexterity may still be developing in the boys) The study was not so planned as to throw light on these various alternatives Seashore (696a) has shown that college men do substantially better than the norms Perhaps in the future more persons planning test research will recognize the futility of merely compiling normative data for relatively undescribed groups, and so set up their research as to provide for answers to questions such as these As in the case of the Minnesota Clerical Test, the applicability of this test to adolescents is still in doubt

Content The test consists of a formboard in which are four rows of identical holes, with fifteen holes in each row Sixty identical discs, each somewhat larger than a checker, fit into these holes, the thickness of the discs being greater than that of the formboard so that they may be readily grasped while in place The flat sides of the discs are differently painted, so that they contrast with the board and so that a ready check may be made in the Turning Test (Educational Test Bureau form only) This test consists of administering the test with the discs in place, but to be turned over and returned to their places by the examinee, the Placing Test (both forms) involves moving the discs from the table-top to the holes in the formboard

Administration and Scoring The test is administered individually, with the subject standing at a table of normal height The examiner places the board with discs on his own side of the table, leaving a little more room between the board and the examinee's edge than is required to accommodate the board The formboard is then raised, leaving the discs on the table and undisturbed The formboard is then placed between the discs and the examinee, about one inch from the edge of the table All this is as recommended in the manuals, administration is further simplified if the psychometrist uses a light board or tray open on one side as a base for the formboard, sliding the latter off the base or tray

to place the discs and putting the base back under the formboard when placing it in front of the subject for testing This makes it possible to lift and remove the formboard without losing discs, and has them in place for the next administration The test is administered in four trials, requiring from six to eight minutes all told

The scoring used in the original MESRI studies added all four trials, Darley (187) has shown, however, that greater reliability is achieved by using the first trial as practice and adding the time required in the last three as the score The revised Minnesota manual gives appropriate norms

Variations have been tried also by Jurgensen (413) and Wilson (932) In the former study Jurgensen used nine methods of administration, some involving use of one hand, some the other, and some both (When both hands are used, blocks are picked up from the same *row,* in adjacent *columns,* except in the last, odd, column) Although he concluded that his revision is more valid and more reliable, and that the part scores are more independent than in the standard version, this method has not been widely taken up It nevertheless merits consideration, along with other variations, when the test is to be validated as part of an employee-selection program, for some variations will almost certainly be more valid for some jobs and less so for others, because of the operation of specific factors Wilson's modification consisted of using only the lowest of three trials rather than the total time, but he gives opinion rather than evidence, convenience rather than validity, as justification for the procedure

Norms As was previously indicated, norms for the University of Minnesota form are available, for the Placing Test only, for the MESRI standard sample of 500 adult workers, and for about a dozen occupations such as butter-wrappers, food-packers, bank tellers, typists, and garage mechanics, represented by from 14 to 164 persons each Although these small occupational groups were sufficiently large to supply answers to some of the questions studied at the Minnesota Institute, and are more varied than those upon which most aptitude tests antedating World War II were based, they are not satisfactory for vocational guidance or selection Data based on them throw a great deal of light on the nature of the traits tested, but it is altogether possible that norms based on larger and more representative groups would differ considerably from these The best test-construction projects of the war and post-war era have recognized the need for larger as well as more varied norm groups as these projects are completed and become better known the better norming of tests such as

this will become a necessity, from a marketing as well as from a professional point of view It might be added, parenthetically, that this relatively new recognition of the need for large-scale occupational norms is virtually taking test construction out of the hands of individual psychologists, who will still originate test ideas, and is putting it in the hands of consulting organizations and test publishers who have the financial resources to subsidize the extensive standardization research which must precede publication It will also take test publication out of the hands of publishers who merely print and sell tests without carrying on or subsidizing test standardization

The Educational Test Bureau form supplies norms for both Placing and Turning Tests, and for three additional variations developed by Jurgensen (413), as pointed out above, MESRI norms which are included in the Bureau norms in some unspecified manner should not be used with this different form until evidence is produced to show that the difference in the formboards does not affect performance Subsequent studies by Teegarden (815,816) Tuckman (877), Jurgensen (413), Seashore (696a) and Cook and Barre (170) have used this form, and make available other sets of norms Teegarden's are perhaps the most useful, for she sampled a dozen jobs represented by applicants at the Cincinnati Employment Center, a white group ranging in age from 16 to 25 As they were a young group, their experience was somewhat limited and their occupations in many cases as yet unsettled In her first two papers (815) Teegarden gives norms for this group of 500 young men and 360 women taken as a group in the last paper (816) she gives data on occupational differences The fields represented include such entry jobs as helpers in skilled trades, operatives of factory machines, factory operatives (hand), packers and wrappers, restaurant workers, and assemblers, inspectors, and testers, together with more adult occupations such as manual laborers, truck drivers and chauffeurs, and sales clerks The numbers in these occupational groups, as in the MESRI samples, were small, ranging from 26 (truck loaders and helpers) to 123 (women domestics) Like the MESRI norms, they give one an understanding of the test and of the significance of arm-and-hand dexterity in various types of work (topic dealt with below), but they are neither large enough nor well-enough selected to serve as norms in the usual sense of the word

Tuckman's norms are for 1117 subjects aged 18 to 58, tested at the Jewish Vocational Service in Cleveland This group was interested in all

types of work, and had varying amounts of education and mental ability, but as clients of a guidance and placement office they were not representative of adults-in-general 365 were high school students, 407 adult men, and 345 women, and the mean age was 22 The Cleveland boys' and girls Placing norms were approximately the same as those provided by the Educational Test Bureau, but the adults were faster than the original norm group, on the Turning Test, all of the Cleveland groups were faster Men excelled most in Placing

Jurgensen tested 212 male paper-mill operatives aged 18 to 31 These norms were combined with MESRI and other data in a way not indicated by the 1946 manual Seashore's data are for two groups of 96 and 48 college men They did much better than the norm group

Cook and Barre tested 468 men and 2007 women applicants for manufacturing employment providing new norms for 18 to 25 year-olds This group differs from Teegarden's and Tuckman's in that it was a factory population, at least temporarily, Teegarden's subjects were willing to accept "anything," but some were clerical, sales, and service workers by background, and Tuckman's included many from the professional and managerial levels, the median Otis percentile being 74 for men and 52 for women As might be expected under the circumstances, Cook and Barre's norms differ from the old, being higher Like Tuckman, they found that sex norms were needed for the Placing, but not for the Turning Test The writer is inclined to believe that Teegarden's norms are the most helpful to the user of the Educational Test Bureau form in counseling since, like the MESRI norms for the university form, they make possible some differential interpretation but they should not, for reasons given above, be used mechanically In selection, local norms should be developed, using the available occupational norms only as a source of ideas as to situations in which the test may prove useful

Standardization and Initial Validation The original work with the Minnesota Manual Dexterity Test, apart from the study of its role in spatial visualization tests, having been carried out as part of the operations of the Employment Stabilization Research Institute, the standardization and initial validation data have to do only with the reliability and occupational differentiation of the test Its reliability is taken up below Its ability to differentiate workers in various occupations was demonstrated by the occupational norms discussed above Highest scores were made by women butter packers and wrappers and by women food packers,

who stood at the 94th, 92nd, and 88th percentiles respectively, while semi-skilled workers in general stood at the 64th Apparently arm-and-hand dexterity is important especially in packing and wrapping jobs

Bank tellers were at the 85th percentile, ranking highest among clerical workers, with men office clerks at the 77th, which suggests that, although there is no correlation between manual dexterity and success in office work, office workers are a somewhat select group in dexterity Since the 164 women office clerks were only at the 60th percentile when compared to general adults, and since the male office clerks were only 66 in number, this may be partly a result of sampling It is probably wise to suspend judgment concerning the importance of manual dexterity in clerical work, operating on the conclusion that the critical minimum is rather low, a conclusion which is in accord at least with the data concerning women clerks

Finally, it should be noted that the skilled groups tested in the Minnesota project did not differ greatly from the mean of the general adult population Skilled workers in general averaged at the 60th percentile, while garage mechanics, to cite a specific example, were at the 55th when compared to employed adults This bears out the statement made at the beginning of this chapter, to the effect that skilled workers depend not on their manual skill, but on other aptitudes and upon technical knowledge

Reliability Darley reported that the reliability of the Placing Test was above 90 for the standard sample (187) Tuckman also used the odd-even method (878), reporting corrected reliabilities of more than 90 for his samples He obtained retest reliability coefficients which were slightly lower, probably because of practice effect In this connection, he confirmed Darley's finding that it is best to use the first trial for practice, the mean score for the first trial was at the 52nd percentile, while that for the 5th trial was at the 79th, a substantial improvement Jurgensen (413) found reliabilities of 87 and 91 for 212 adult men

Validity The first step to be taken in ascertaining the validity of the Minnesota Rate of Manipulation Test would seem to be to determine how *independent* the two parts are This was done by Blum (106), Jacobsen (396), Jurgensen (413), Seashore (696a), Teegarden (815) and Tuckman (878) The first obtained a correlation of 55 based on 120 women packers and wrappers, which compares very favorably with that of 57 reported in the test manual Jacobsen's intercorrelation was only 27, for 90 aircraft industry trainees Jurgensen's intercorrelation was 52 for 212 adult male

paper-mill operatives Seashore reported correlations of 46 and 58 for two samples of college men Tuckman's figures were higher, being 60 for 345 women and 66 for 407 men Teegarden's were still higher, at 65 for 171 women and 73 for 230 men Presumably the true correlation is about 60 for women, and somewhat higher for men (only Jacobsen's study is out of line), indicating that the two tests are measuring the same basic aptitude manifesting itself in two slightly different ways, or that they have an important factor in common but one or more others peculiar to one and not to the other The factor analysis study carried out by the United States Employment Service (735), referred to early in this chapter, showed that the former hypothesis is correct, and that the Placing and Turning Tests have practically identical factor composition, they are almost pure tests of arm-and-hand dexterity

The Manual Dexterity Test has been correlated with tests of intelligence by Tuckman (878), Jacobsen (396), and Super (unpublished study). Tuckman administered the A C E Psychological Examination to high school students and adults, finding correlations of 18 and 17 for Placing, and 29 and 26 for Turning Job analysis of the tests suggests that the closer relationship between Turning and intelligence may be due to the slightly more complex manual task in that test which requires bimanual co-ordination of a rudimentary sort But Jacobsen found correlations of 16 and 12, using 90 adult subjects Administering the Otis S A Test to 100 NYA youth, the writer obtained a correlation of 11 with the Placing Test In any case, the role of intelligence is negligible

The similarity of fine-manual to gross-manual dexterity was ascertained by Roberts (633), Jacobsen (396) and by Blum and Candee (105) Roberts (633) found correlations of 46 and 40 between Placing and Turning, on the one hand, and his Pennsylvania Bimanual Worksample (Assembly) Test, a nut-and-bolt assembly task somewhat finer than the Minnesota Test but grosser in its requirements than the O'Connor Tests (N = 473) Jacobsen tested 90 wartime aircraft industry trainees, and found correlations of 20 and 06 between O'Connor Finger Dexterity and Placing and Turning Tests, of 26 and 20 between Tweezer Dexterity and the two Minnesota subtests Only the highest of these correlations was statistically significant Blum and Candee tested 130 women packers and wrappers in a department store with the O'Connor Finger Dexterity Test, finding correlations of 42 and 335 with Placing and Turning Tests The correlations were reliable With only two studies available, one with negative findings and one with positive, we are faced with a dilemma But poor

testing conditions and other defects of procedure are more likely to produce negative findings than positive, and the negative study was the work of a beginner while the positive was that of two experienced investigators It therefore seems necessary tentatively to conclude that there is some relationship between arm-and-hand dexterity on the one hand, and wrist-and-finger dexterity on the other That the relationship is not high is indicated not only by these data, but also by the USES factor analysis (735) which isolated two relatively independent manual factors one gross and one fine

The role of arm-and-hand dexterity in tests of *mechanical comprehension* was studied by Jacobsen (396) and Super (unpublished study) The latter found a correlation of only 05 between Minnesota Mechanical Assembly Test scores and the Placing Test, the subjects being 100 boys and girls aged 16 to 24 employed on NYA projects This is noteworthy, as the Assembly test involves the putting together of a variety of mechanical objects such as a spark plug, a mechanical bottle-stopper, and an old-fashioned lock It is confirmed by Harrell's (336) factor analysis of the Minnesota Mechanical and other tests, which showed no manual dexterity factor in the Minnesota Mechanical Assembly Test Jacobsen found correlations of 21 and 14 between Placing and Turning, on the one hand, and the Bennett Mechanical Comprehension Test on the other The latter is a paper-and-pencil test measuring a somewhat higher order of mechanical comprehension than the assembly test, that it has no significant relationship to manual dexterity is therefore not surprising

Tests of spatial visualization have been correlated with the Minnesota Rate of Manipulation Test by Jacobsen (396), Teegarden (815) and Super (unpublished study) The Minnesota Paper Form Board had correlations of 06 and 00 with Placing and Turning in Jacobsen's study, as compared with one of 23 with the Placing Test in the writer's investigation The writer found a correlation of 05 between the Minnesota Spatial Relations Test and the Placing Test, a relationship which he has never seen reported in the literature although it was to compute it that Ziegler constructed the latter test Jacobsen supplied the correlation between the Crawford Spatial Relations Test and the Placing and Turning Tests 19 and 11 As none of the above relationships were statistically significant it is clear that manual dexterity and spatial visualization tests are independent

Ratings on success in training were used as a criterion in only one published study with the Minnesota Rate of Manipulation Test This was

Jacobsen's investigation of the relationship between success in training in aircraft mechanics and scores on various aptitude tests (396) These war-industry trainees were rated by their instructors after the first two weeks of training, and periodically each month thereafter for the two or three months of training Ratings were for seven traits such as learning, speed and co-ordination, workmanship, and personal fitness for the occupation, rated on a five-point scale As the specific traits had correlations with total fitness ratings which ranged from 84 to 97 the latter only were used as a criterion, all fitness ratings for a given individual were combined, apparently no attempt was made to ascertain the reliability of the ratings, although the data would have permitted it Correlations ranged from — 03 to 17, none of them being reliable Either gross manual dexterity as manifested in aircraft engine, aero repair, machine shop, and other similar courses has no bearing on instructors' evaluations of mechanical promise, even though they rated the subjects for speed and co-ordination, or the Minnesota Rate of Manipulation Test does not measure the type of manual dexterity sought by instructors As will be seen later, fine-manual dexterity as measured by the O'Connor tests had unreliable and low correlations with these same ratings, the only tests which give reliable predictions of instructors' ratings in these courses being mechanical comprehension, arithmetic, and intelligence tests, in that order

Success on the job has been studied with electrical worksamples, production in department-store packing and wrapping, supervisors' ratings of efficiency on these same jobs, ratings of efficiency in pharmaceutical inspecting and packing, and ratings of success in ordnance factory and paper-mill employees

The electrical *worksample* developed by O'Rourke (connecting a push-button, bell, and dry-cell) was used with 49 boys and 37 girls aged about 18, employed on NYA projects, by Steel, Balinsky, and Long (751) Tests were administered to one-half of the subjects before the projects were initiated and to the others after the worksample The worksample was carried out individually in order to permit careful observation by the examiners, who recorded care in the use of directions, facility in handling tools and materials, initial adjustment to the task, and reaction to difficulties Brief interviews were held after the project in order to elicit further reactions, but only the time score on the worksample was used in the correlations Neither of these was statistically significant for boys (— 02 and 10 for Placing and Turning), but both were for girls (50 and

35) Other data throw light on the reasons for these discrepancies The boys took significantly less time to complete the worksample than the girls, although this was not true of the tests, the boys had had more experience with electrical equipment and tools Apparently it was amount of experience with electrical equipment which determined the boys' time scores, rather than gross-manual dexterity or any of the abilities measured by the other tests (fine-manual dexterity, spatial visualization, and vocabulary), but unfortunately no test of electrical information was used to provide a quantitative check on this explanation The girls, on the other hand, had had so little experience with such equipment that vocabulary (ability to understand and follow the directions), spatial visualization, and both types of manual dexterity determined the amount of time they required to complete the task Mechanical comprehension, had it been tested, might also have played a part in the case of the girls, since the other relevant aptitudes were important to their success These conclusions suggest that manual dexterity and other aptitude tests are most likely to be valuable when selecting inexperienced workers for semiskilled jobs, or for counseling inexperienced persons, whereas careful evaluations of experience are likely to be more valuable with those who have had relevant experience This conclusion applies only to *initial job adjustments* in semiskilled work, however, for that is what the worksample tested, as Blum and Candee's study of packers and wrappers (106) showed, skills that are important in initial adjustment sometimes play no part in long-term success, other aptitudes emerging as the important ones after some experience has been acquired

In their first study of department store *packers and wrappers* Blum and Candee (105) tested 38 permanent employees of one department store, together with 52 employment-service applicants subsequently employed by the store for whom criterion data became available The criteria used were production records and supervisors' ratings For the former, the average daily number of packages wrapped during the month of December, when employees work most nearly at their capacity, was used, its split-half reliability was 88 The supervisors' ratings were those routinely made on a four-point scale, and consisted of an overall efficiency rating phrased in terms of recommended continued employment and seasonal rehiring No data are presented concerning the reliability of the ratings, which included none in the "inefficient" or lowest category

The correlations between production records and Placing and Turn-

ing Tests were 35 and 27 for seasonal employees, and 21 and 06 for permanent employees Evidently arm and-hand dexterity plays some part in initial job adjustment in packing, but the skill requirements are actually low enough so that experience erases the effect of differences in aptitude When supervisors' ratings were used as a criterion no significant differences were found between superior and inferior seasonal nor between superior and inferior permanent employees, although the permanent employees were rated superior to the seasonal employees and made higher test scores As the seasonal employees were considered especially good that year, although not actually superior to the general population, Blum and Candee concluded that experience must affect test performance While this is perhaps true, it cannot be considered as having been proved, for there was no pre-employment testing of the experienced group and no post-employment testing of the inexperienced group, the higher scores of the experienced group may have been due to self-selection, through the quitting of satisfactory experienced workers who found that relative lack of manual dexterity required them to put forth a disproportionate and unsatisfactory amount of effort in order to keep up

In their second study (106), Blum and Candee tested comparable groups in another department store, and used similar criteria, but the Turning Test was omitted Again there was a moderate but significant relationship between arm-and-hand dexterity and production in the case of packers who handle large items (37), but not in the case of wrappers who handle small items and make change Again there was no relationship between test scores and supervisors' ratings of permanent employees, but seasonal employees given the highest ratings tended to make slightly higher test scores than those given lower ratings The general conclusion is the same as for the first study arm-and-hand dexterity plays a part in the initial job adjustment of packers, whose movements are gross in nature, but practice minimizes its effects, in the case of wrappers, whose work involves somewhat finer but still gross movements, arm-and-hand dexterity as measured by the Minnesota test played no part Neither did finger dexterity as measured by O'Connor's test, perhaps a new test of an intermediate degree of fineness, involving wrist-and-finger movements with objects the size of the Minnesota Rate of Manipulation Test's discs, would have produced positive results

Another study of the same type was made by Ghiselli (288) with 42 seasonal wrappers who were rated for both quality and quantity of

output The ratings were combined, and proved to have no significant relationship to Placing and Turning Test scores (— 10 and — 02) Data for finger dexterity were approximately the same Ghiselli (286) also worked with pharmaceutical inspector-packers, whose tasks consisted of filling, stoppering, examining, labeling, cartoning, and packaging containers of fluids, powders, and pastes Job analysis suggested that arm-and-hand dexterity and eye-hand co-ordination should be among the important characteristics in performing the work The Minnesota Rate of Manipulation Test was therefore among those included in the battery Production being difficult to measure because of variations in the nature of the work, a rating scale was devised to measure the traits suggested by the job analysis, and the forewoman in charge of the work rated each of the 26 girls In addition, the supervisor of the finishing room rated each on overall value to the organization Reliability of the ratings was checked by correlating the composite forewoman's ratings with the supervisor's overall rating the coefficient was 72 The two ratings were therefore combined to serve as criterion The correlations between criterion and Placing and Turning Tests were — 24 and — 40 (negative because the scores are in terms of seconds used to perform the task) Of the other factors measured spatial visualization was the most important, more so than manual dexterity, clerical perception was important, but less so than manual dexterity, and some of the spatial and eye-hand co-ordination parts of the MacQuarrie were also valid Ghiselli s preliminary job analysis therefore proved to be a sound one It is interesting that the specific factors in the Turning Test made it more valid than the Placing, even though they measure the same basic factor further evidence of the desirability of using custom-built test batteries, and even custom-built tests, for selection purposes

It is also noteworthy that, although the manual operations in the pharmaceutical job appear to have been more like those of the wrappers than those of the packers in Blum and Candee's studies, the dexterity test had less validity for wrapper selection than for packer selection, and less for packers than for pharmaceutical inspector-packers In Blum and Candee's studies manual dexterity had some predictive value for initial job adjustment, but no validity for experienced workers, while in Ghiselli's pharmaceutical study no distinction on the basis of experience was made Herein perhaps lies the explanation of the apparent discrepancy Ghiselli's group is not described in terms of specific experience, but the general statement is made that both the rate of turnover in the

department and company morale were high If the group contained as great a range of experience as these facts suggest, and if manual dexterity is a selective factor on the job, then the range of manual dexterity was probably greater in Ghiselli's sample than in Blum and Candee's (the published data do not permit comparisons) A difference in aptitude sampling such as this would result in a higher correlation coefficient for Ghiselli's study, even though the role of manual dexterity were really identical in the two occupations The final judgment would seem to be that Blum and Candee's conclusions are correct, but that the role of manual dexterity is somewhat greater than they found it to be

A final study of the role of manual dexterity in packing jobs is that carried out by the United States Employment Service and cited by Stead and Shartle (750 217–227) They administered the Minnesota test to 43 can packers, 30 merchandise packers, and 41 inspector-wrappers (the jobs are not further described) A production criterion was used for the first two jobs average number of cans packed per hour, and ratio of time estimated as needed to complete a unit of work by time-and-motion study men to time actually used to complete the unit, a rating was used for the last-named job The correlations for the Placing Test were 35, 14, and — 09, for the Turning Test, 22, 11, and 01 respectively Only for the can packers are the correlations high enough to be significant, and the relationship is the opposite of that anticipated (true also for finger dexterity) the slower or less dextrous tended to have the greater output For the inspector-packers, whom Ghiselli had considered most likely to resemble his group, no relationship was found The merchandise packers closely resembled Blum and Candee's department store packers in operations performed the correlations are lower in this study than in theirs Failing more detailed data on the USES study, reconciliation of its findings with the others seems impossible if enough facts were available, good reasons for the discrepancy would no doubt be found Perhaps the USES study merely reversed signs It therefore seems wise to abide by the conclusions drawn from the studies which have been reported in more detail The USES study also included pull-socket assemblers, put-in-coil girls, and cafeteria counter and floor girls, for none of whom the test had validity (r's = — 15 to 19)

A different type of occupational group was studied by McMurry and Johnson (500), who administered the Minnesota dexterity test to 768 women being hired by an ordnance factory Scores were validated against ratings of 587 who remained long enough to be rated The reliability of

the ratings was apparently not checked Distribution among jobs is illustrated by the fact that there were 97 welders, 140 assembly workers, and 33 inspectors No validities were reported, however, for the Rate of Manipulation Test

The *paper-mill employees* studied by Jurgensen (413) were men hired as converting-machine operators, whose work consisted mostly of removing a specified number of tissue-paper sheets from the machine, raising the top sheet to insert advertising material, and placing the package of sheets on a conveyor All 60 were right-handed high school graduates between the ages of 18 and 31 The criterion was a combination of three supervisors' ratings, the reliability of which was 75 Placing and Turning Tests were both administered, plus some variations which included placing and turning with both hands simultaneously Validity coefficients were Placing 325, Turning 455, Right-Hand Placing Turning 57, Simultaneous Placing-Turning 33 These findings indicate not only that the Minnesota test has predictive value for this type of semiskilled factory work, but also that motion study can be valuable in suggesting variations in the test which increase its validity for specific operations It is regrettable that Jurgensen did not also utilize an output criterion, the greater objectivity of which (if not affected by slow-down, poor morale, etc) would provide a better index

Occupational differentiation by means of the Minnesota Rate of Manipulation Test has been checked by a number of investigators and for a variety of jobs Blum and Candee (106) found that satisfactory experienced department store packers and wrappers made better scores than the general population on the Placing Test but the seasonal workers who were considered an exceptionally good group did not Ghiselli (286) reported that pharmaceutical inspector-packers stood at the 96th percentile on the Placing and at the 91st on the Turning Tests when compared to the general population In the USES study (750) pull-socket assemblers, put-in-coil girls, and can packers exceeded the 75th percentile of the general adult population in Placing, in Turning the merchandise packers displaced the can packers Merchandise packers, cafeteria counter and floor girls, and inspector-wrappers were in the normal range in Placing, with can packers taking the place of the merchandise packers in Turning Teegarden provided data for groups of from 26 to 123 semiskilled workers in a study previously cited Of these occupational groups only the assemblers, inspectors, and testers stood at about the third quartile, with women packers and wrappers slightly below it, on

both Placing and Turning Tests, male packers and wrappers, helpers in skilled trades, factory hand operatives, machine operators, and women clerks were about one sigma above the adult mean in Placing, the women machine operators and packers and wrappers being there also in Turning, truck drivers and chauffeurs, truck loaders and helpers, male sales clerks, restaurant workers, domestic workers, and manual laborers all being in the normal range It has already been noted that in the MESRI studies butter and food packers and wrappers, bank tellers, and male office clerks ranked higher than the third quartile on the Placing Test, while women office clerks, minor bank officials, minor executives, semiskilled workers, stenographers, typists, and garage mechanics, were in the average range From these findings it seems legitimate to conclude that arm and-hand dexterity as measured by the Minnesota test is important in packing, wrapping and inspection jobs and in gross-manual assembly and machine-operation jobs, the predictive value of the test depends somewhat, however, upon the specific factors in the job and the degree to which they also are tapped by the test For this reason sometimes the Placing Test, sometimes the Turning Test, and sometimes other variations such as those tried by Jurgensen (113) with the Minnesota materials and by others with custom-built pegboards, will have the most predictive value and so be most helpful in selection or counseling

Job satisfaction, in the case of the Minnesota Manual Dexterity Test as in that of the Minnesota Clerical Test, has apparently not been a subject of investigation

Use of the Minnesota Rate of Manipulation Test The Minnesota Manual Dexterity or Rate of Manipulation Test has been found to be useful primarily in connection with semiskilled occupations in which skill in arm-and-hand movements seems, in job analyses, to be important It has not been found valuable in skilled trades, in which understanding of the processes involved is more important than individual differences in the manual dexterity with which they are executed. Even in the grosser manual jobs such as packing and the assembly of large parts, differences in skill which are found to exist before employment play a part primarily in initial adjustments to the work rather than in long-term adjustments, practice in the specific job operations appears to reduce the effect of preemployment differences to the zero point It may be that these differences play a part at this stage which current studies have not brought out, by making the maintenance of adequate production so easy as to render the work satisfying, or such a strain that it becomes subtly unbearable and

makes the worker quit or continue in a state of undiagnosed dissatisfaction In the light of present knowledge, however, this test seems likely to be useful in counseling inexperienced persons concerning the choice of packing and assembly jobs It is even more likely to find use in selection, when quick adjustment to routine work is desired, than in counseling

In *selection* programs local norms should be used, and in the initial studies of the test with a given job in a given plant variations in the technique of the test-task should be tried The test then taps specific factors in the job along with the basic or group factor which should be its principal source of validity The nature of these variations is suggested by job analysis The validity of the test is increased by this method, for the initial job-adjustment period, its long-term adjustment validity may be decreased by the emphasis on developed rather than on latent skills, but at this stage of our knowledge that is only a subject for speculation and investigation

It is doubtful whether this type of test has any place as a directional instrument in a *school counseling* program If experience erases the effects of normal individual differences in this type of dexterity, then it is the function of education to provide such experience in appropriate cases (those of persons who may enter such work, as suggested by intelligence, interest, and socio-economic status) The test will not be useful in providing data for the making of decisions concerning the choice of semiskilled occupations It may, on the other hand, give some insight into the assets and liabilities with which a student enters upon new experiences

In *employment counseling*, whether at the end of an educational program, in an adult guidance center, or in an employment service, the test should be of more value, for the question of initial job adjustments of workers inexperienced in jobs requiring arm and-hand dexterity is both more common and one on which the Minnesota test throws some light Since the occupational norms are based on small groups they must be selected with a full understanding of the particular sample and employed tentatively, but some facts cautiously used are better than none at all when decisions have to be made When the University of Minnesota's form is used, the MESRI norms are still the best, when the Educational Test Bureau's form is used, then Teegarden's data will probably be found most helpful In either case the norms should be thought of as merely suggestive

Finally, it is pertinent to ask whether both Placing and Turning Tests should be used, or only one of them, and if the latter, which one In a specialized battery for semiskilled jobs both should be used, because of the higher correlation of the Placing Test with some gross-movement jobs such as department-store packing, and of the superior validity of the Turning Test for some finer-movement jobs such as packaging drugs In more comprehensive test batteries, in which there is not sufficient testing time for the refined investigation of each area, the fact that the factor composition of the two tests is identical means that one of the tests should be sufficient for survey or screening purposes In such situations the test to be used should depend upon which is likely to be more closely related to jobs the examinee may consider or be considered for, in the absence of data for the making of such judgments, the Placing Test can probably best be used, together with a wrist-and-finger dexterity test to tap the other extreme of fineness

The O'Connor Finger and Tweezer Dexterity Tests (C H Stoelting and Co , 1928)

The O'Connor Finger and Tweezer Dexterity Tests were developed in the middle 1920's while O'Connor was employed in the West Lynn works of the General Electric Company (374) He was concerned with the selection of women for electric-meter and instrument assembly work, and devised these tests for that purpose Similar tests had previously been described by Whitman (923), who used them with children They have since been tried out on various types of workers, particularly in the Minnesota Employment Stabilization Research Institute

Applicability The tests were designed for use with adults and with older adolescents of post-high school age, they were standardized on such groups, and restandardized on similar groups by the MESRI project They are widely administered to adolescents, but this writer has seen no studies of their applicability to these younger groups The fact that physical maturity comes somewhat earlier than mental has seemed to warrant the use of this dexterity test from age 13 or 14 on (94), but it has not actually been demonstrated that this specific type of dexterity matures early We have seen that the assumption of early maturation proved misleading in the case of the Minnesota clerical and manual dexterity tests it may be equally misleading in the case of this instrument In the absence of data on this question, one should proceed cau-

tiously with the use of the O'Connor tests with high school boys and girls but there is probably not much danger of being misled as a result of age-changes after the last two years of high school

Candee and Blum (133) have reported that age in adulthood and work experience have no effect on the scores of the O'Connor tests

Content The Finger Dexterity Test consists of a shallow tray beside a metal plate in which there are 100 holes arranged in ten rows of ten holes each (the only readily available form, Stoelting's, is made of different materials) Each hole is large enough to hold three metal pins one inch long and .07 inches in diameter, the holes are spaced one-half inch apart The Tweezer Dexterity Test is sometimes the opposite side of the boards used for the Finger Dexterity Test, again the metal plate has 100 holes in it, but these are only slightly larger than the pins, allowing one to be placed in each hole A pair of 00 gauge tweezers are used in this test to pick up the pins

Administration and Scoring The Dexterity Test is administered with the subject seated at a table of standard height (30 inches), with the pinboard about a foot from the edge of the table, the tray on the side of the favored hand, placed at an angle of about 90 degrees to the subject The directions are clear except for one point The O'Connor tests are incorrectly given by many psychometrists because they do not read the instructions carefully enough to realize that if a right-handed subject were to start in the top left corner and fill the holes toward himself he would fill *columns*, which go *vertically*, up-and-down, rather than *rows* The examinee should actually begin at the far corner (top-left for a right-handed person) and fill the holes of the top *row across* to the other (top-right for a right-handed person) corner, then begin to fill the holes in the second row in the same manner as the first, then the holes of the third row, etc

In the Finger Dexterity Test the subject picks up three pins with his preferred hand and places them in each hole, in the Tweezer Dexterity Test he picks up one pin at a time and places it in its hole The score is normally on the basis of time, with a small correction of the second half for practice on the first half, some recent studies, including those of the USES, use simply the total time required, which is probably sufficiently refined for practical purposes The time required varies from 8 to 15 minutes for the Finger Dexterity Test, and up to about 10 minutes for the Tweezer Test Accurate timing is important and requires either a stop watch or a watch with a sweep-second hand

Norms Although O'Connor presented adult norms in his original report of his work (374), the most representative and generally used norms are those of the Minnesota Employment Stabilization Research Institute (306) These are for the standard sample of 500 employed adults, supplemented by averages for small groups of persons in a variety of occupations, most of which are unfortunately not the types for which these tests can be expected to be useful Means and sigmas are available for more pertinent occupations in other studies (103,750) discussed below, but unfortunately the scores in these are given in terms of total number of seconds used rather than in terms of O'Connor's correction Perhaps in due course the work of the USES will make it advisable to use the total time score and their norms, in the meantime, the corrected score and the MESRI norms are best As no published manual is available, the Minnesota norms are reproduced in Table 14

TABLE 14

NORMS FOR THE O'CONNOR FINGER AND TWEEZER DEXTERITY TESTS, MESRI STANDARD SAMPLE EMPLOYED ADULTS

Raw Score Men		*Raw Score Women*			
F D	*T D*	*F D*	*T D*	*Standard Score*	*Percentile*
183	—	166	—	8 0	99 9
194	255	175	249	7 5	99 4
207	271	186	263	7 0	97 7
221	289	197	279	6 5	93 3
238	309	211	297	6 0	84 1
257	333	226	318	5 5	69 1
280	360	244	342	5 0	50 0
307	393	265	369	4 5	30 9
340	432	290	401	4 0	15 9
382	479	319	440	3 5	6 7
434	539	356	487	3 0	2 3
503	615	402	544	2 5	6
598	—	462	—	2 0	14

Means for occupational groups which might be expected to make high scores on these tests, together with those of certain others included for sake of contrast, are given in Table 15 and can serve as a suggestive guide in the use of O'Connor test results

It should be noted that women tend to do better than men on this as on other types of dexterity tests, and that the only occupations for which these tests have been shown to have any clear-cut value are women instrument assemblers, bank tellers, office workers, manual-training teachers, and draftsmen, as will be brought out below, these data help one to understand the test, but they are hardly enough to serve as norms

TABLE 15

AVERAGE FINGER AND TWEEZER DEXTERITY SCORES OF SELECTED OCCUPATIONAL GROUPS

No	*Sex*	*Occupation*	*F D Mean Score*	*F D Md Centile*	*T D Mean Score*	*T D Md Centile*
17	M	Bank Tellers	243	80	325	76
113	M	Office Clerks	255	70	323	76
170	M	Manual Training Teachers	258	67	327	74
21	M	Draftsmen	259	69	335	70
61	M	Ornamental Iron Workers	271	57	341	65
102	M	Garage Mechanics	278	51	352	56
31	M	Machine Operators	331	18	385	34
228	M	Casual Laborers	385	7	570 5	1
*	F	Instrument Assemblers	219	76	*	*
180	F	Stenographers-Typists	230	65	333	57
21	F	Office-Mach Operators	231	64	*	*
15	F	Food Packers	235	59	345	48
19	F	Butter Packers	248	47	340	50
317	F	Graduate Nurses	252	42	334	55

* Data not available

Standardization and Initial Validation O'Connor standardized the test on 2000 women applicants for factory employment and an equal number of men in the General Electric plant at West Lynn, Massachusetts The Finger Dexterity Test was administered a number of times to the same workers, with the finding that the second trial was somewhat better than the first, and the fifth trial showed little further improvement over the fourth Retest reliability for the first and second trials was 60 on the Finger Dexterity Test, considerably lower than those obtained by others The original validation of this test was on a group of 36 women applicants who were tested when interviewed for employment and hired for assembly work Hines and O'Connor (374) reported that 36 percent of those in the lowest quarter left the company before 8 months had elapsed, as compared with only 6 percent of those in the top quarter, this seems impressive until it is realized that 36 percent of one-quarter of 36 (the total number of cases) is slightly more than one-third of 9, that is, three-and-a-fraction persons, and that 6 percent of one-quarter of 36 is approximately $\frac{1}{17}$ of 9, or a little more than one-half of one person Just how three-and-a-fraction persons, and a little more than one-half a person, can fail is something that even some eminent test-construction specialists have failed to ask (590 237) The report does not exactly strengthen one's confidence in the original work with the test,

however ingenious the test idea No data were published by O'Connor dealing in specific detail with the Tweezer Dexterity Test

Reliability As we have seen, Hines and O'Connor (374) originally reported the retest reliability of the Finger Dexterity Test as .60 Blum (103) retested 64 employment applicants, obtaining a much higher coefficient of 89, he also reported an uncorrected split-half reliability of 77. Split-half reliabilities for the same test have been reported by Darley (187) these are (corrected) 93 and 90 for samples of 475 men and 215 women Apparently the test is reliable even according to the retest method No reliability data have been located for the Tweezer Dexterity Test, the above investigators having made the seemingly warranted assumption that the two tests cannot differ much in this respect

Validity Because of their early publication the O'Connor dexterity tests have been used in a number of studies Even though many of these had only an indirect or very partial interest in the nature and validity of the test, they do, taken as a whole, throw considerable light on its validity

Correlations with other tests have been computed for the usual variety of measures The Finger and Tweezer Dexterity Tests have been found to have intercorrelations of 17 by Jacobsen (396) with 90 war-industry trainees as subjects, 19 by Blum (103) who tested 119 women factory-employment applicants, 47 by Thompson (824) with 35 dental freshmen, 33 by the Minnesota project (187) with a heterogeneous group of women and 56 for a similar group of men, and 57 by Harris (341) with a group of 66 dental students 59 of whom completed the four year course As Blum's and Jacobsen's results are based on factory workers, the others' on professional or mixed groups, it is probably safe to conclude that the correlation is approximately 50 in heterogeneous groups and less than 20 in homogeneous groups

Correlations between O'Connor dexterity tests and *intelligence* tests have been reported by Harris (341) for dental students, using the Otis S A Test The coefficients are — 01 and 015

The relationship with *arm and-hand dexterity* is perhaps of most interest Finger Dexterity was found to correlate to the extent of 21 and 42 with the Minnesota Placing Test by Jacobsen (396) and by Blum and Candee (105), 06 and 335 with the Turning Test For Tweezer Dexterity correlations of 26 and 20 with Placing and Turning Tests were reported by Jacobsen (396) With one exception, Jacobsen's

correlations are not high enough to be significant, while Blum and Candee's are As was brought out in the discussion of arm-and-hand dexterity, it seems likely that the latter's results should be accepted until more conclusive studies are made The two types of dexterity should be thought of as related but distinct aptitudes

Correlations with tests of *spatial visualization* are more numerous, Andrew (21) reported correlations of 28 and 31 between the Minnesota Spatial Relations Test and the Finger and Tweezer Dexterity Tests, based on 200 women clerical workers For the Revised Minnesota Paper Form Board Jacobsen (396) and Thompson (824) reported correlations of less than 15 with war-industry trainees and dental students Jacobsen also found no significant relationships with the Crawford Spatial Relations Test (22 and 11), in a more heterogeneous group they might be higher Harris (311) reported correlations of −02 and 15 with the Wiggly Block Evidently ability to visualize space relations plays no part in wrist-and-finger dexterity

Wrist-and finger dexterity has rarely been correlated with *mechanical comprehension,* probably because of the anticipated low relationship Jacobsen (396) confirmed expectations with coefficients of −08 and 14 with the Bennett Mechanical Comprehension Test

Success in training has been investigated for an electrical worksample, aircraft mechanics, power-sewing-machine operation, machine-tool operation, fine arts, and dentistry

The study of the *electrical worksample* (751) has already been described in connection with the Minnesota Rate of Manipulation Test In it the Finger Dexterity Test had validities of 08 for boys and 35 for girls, while those for the Tweezer Dexterity Test were 18 and 42 (other data show that the signs should be negative, being for time scores) As has already been seen, there is reason for believing that the boys' worksample scores were a reflection of degrees of experience while the girls' were the result of differences in aptitudes, and that in such work wrist-and-finger dexterity is likely to play a part only in the initial adjustments of novices or in output differences in equally experienced workers

The investigation of factors in success in *aircraft mechanic* training was also discussed in the section on the Minnesota Rate of Manipulation Test Jacobsen (396) found only two significant relationships between O Connor dexterity tests and instructors' ratings of fitness for the occupation these were between Finger Dexterity and ratings in aircraft electricity (31) and Tweezer Dexterity and ratings in aircraft instruments

(32) As the other eight coefficients ranged from — 02 to 22, and there is no apparent logic underlying the different results, the writer is inclined to consider the two statistically significant correlations the products of chance In any set of correlation coefficients some will appear significant simply as a result of chance factors Such a conclusion is forced also by the illogic of those who take most time to complete a speed test being the best students (unless Jacobsen reversed signs)

High school girls learning *power-sewing-machine operation* were studied by Otis (579), who used time taken to complete a series of work-samples and quality-ratings of the same tasks as criteria The two criteria had an intercorrelation of — 17± 13, from which it might be concluded that there was no relationship between speed and quality, but which may, in the absence of reliability data, only prove that one or both of the criteria were unreliable The speed criterion had a correlation of 27 with Finger and 46 with Tweezer Dexterity, the quality criterion, 20 and 07, neither of these latter being statistically significant These results suggest that at least the speed criterion was reliable, and show that those who were fastest on the test tended to be most rapid on the task

In the study of *machine-tool trainees* Ross (651) administered the O'Connor Finger Dexterity, Minnesota Spatial Relations, and O'Rourke Mechanical Aptitude Tests and related them to grades in training, establishing critical scores but not obtaining any indices of degrees of relationship

Students of *fine arts* were tested with the O'Connor dexterity tests and the Minnesota Paper Form Board by Thompson (824), the number of students being 50 Correlations with point-hour ratios were 21 for Finger and 08 for Tweezer Dexterity, neither of which was clearly significant This finding can perhaps be discounted, however, since grades in fine arts are probably not the most appropriate of criteria a study using ratings of the quality of artistic craftsmanship, made by experts and checked for reliability, might yield quite different results

Students of *dentistry* have been studied by Douglass and McCullough (208), Harris (341), Jones (unpublished study), and Thompson (824) In the first-named study a variety of tests were tried at the University of Minnesota over a period of several years, with average grades in dental school the criterion The results varied somewhat from one sample to another, but in a typical group of 83 students the correlations between grades and Finger and Tweezer Tests were — 10 and — 30 In Harris'

preliminary study of 50 dental freshmen at Tufts first year grades were the criterion, and the correlations with Finger and Tweezer Dexterity Tests were — 395 and — 36 These are the only studies reporting validity for *grades* in dentistry, it seems rather surprising that so specific an aptitude should show a substantial relationship to such a multi-factorial criterion as college grades, especially as the first two years of dental training are more academic than manual or practical And Harris' more definitive study, in the same school and with the same tests, based on 66 students with both first- and four-year grades as criterion, yielded validities of only — 10 and — 17 for first-year grades and 15 and — 10 for four-year grades (for the numbers in question, the coefficients would need to equal 31 to be significant at the 1 percent level) The Otis S A Test, on the other hand, had validities of 55 and 35 Thompson also correlated O'Connor's tests with freshman and four-year grades, for one group of 35 freshmen and another of 40 seniors in dentistry, finding validities of 01 and 01 for Finger and — 07 and 13 for Tweezer Dexterity E S Jones, in conversation with the writer, has also reported obtaining negligible correlations between O'Connor tests and dental grades at the University of Buffalo The evidence is now very strongly in favor, therefore, of a lack of predictive value in the O'Connor tests for grades in dental school, the statements of O'Connor (571) and the guarded suggestions of Bingham (94 284 and 286) to the contrary notwithstanding However, their logic seems so good that this writer too would not be surprised to see substantial validities for these tests when correlated with a practical criterion, e g , reliable ratings, such as might be made by patients, of skill in clinical work Douglass and McCullough's (208) correlations with laboratory grades, — 43 and — 35, are promising Other studies of this type, and consistent validities, have as yet not been reported, it will be seen that other evidence of the tests' validity for dental training exists

Success on the job has been the subject of investigation with watch assemblers, electrical fixtures and radio assemblers, department store packers and wrappers, pull-socket assemblers, put-in-coil girls, and can packers

In a preliminary study of *watch assemblers* Candee and Blum (133) administered the Finger and Tweezer Dexterity Tests to 20 women workers selected as superior and 17 selected as mediocre by their foremen The difference between scores of the two groups on the Finger

Dexterity Test approached significance (D/σd = 3 18), no such difference was found for the Tweezer Dexterity Test (D/σd = 1 01), but this latter test differentiated employees from a group of applicants better than did the former Critical scores of 7′30″ and 5′30″ were established for Finger and Tweezer Dexterity Tests Two years later these workers were followed up by Blum (103) None of those who had been rated superior had been discharged, as contrasted with 18 percent of the mediocre workers the critical ratio was 2 00 The salary ratios (average weekly piece-rate earnings over a three-month period divided by the average for all employees and expressed as an index with $20 per week equal to 100) of the two groups were 110 and 93, which gave a critical ratio of 3 7, apparently the foremen's judgment of superiority was generally good Although the groups were so small as to make conclusions necessarily tentative, the trend was clearly for superior workers to make better scores on the two tests

In a subsequent study Blum (103) used length of employment, foremen's ratings, and salary ratio as the criterion The salary ratio was that described above, length of employment was divided into "less than one week" (failure group), one week to four months (unsatisfactory group), four months to one year (a moderately proficient group), and more than one year (permanent and effective employees), foremen's ratings were on a five-point scale ranging from "excellent" to "terrible " The first two criteria are objective and hence reliable the last had a reliability coefficient of 60 for 19 workers re-rated after the lapse of more than one year, which is quite high in view of changes in the worker during such a period The subjects were women applicants for factory work at a branch of the New York State Employment Service, 137 constituted the tested group, and another 84 who also were selected solely on the basis of an interview but were not tested were used as a control group Most of the group had had industrial experience but none had worked in watch factories, all were white, and 90 percent were between 20 and 25 years of age, with a range from 18 to 40 The factory at no time had knowledge of the women's test scores The Finger and Tweezer Tests were administered before hiring, scores obtained were time in seconds, quality ratings (reliability for Finger Dexterity equaled 89), and absolute and relative improvement (reliabilities of 13 and 26) It is worth noting that time and quality scores had intercorrelations of 14 for Finger and 71 for Tweezer Dexterity Tests, and that the two quality ratings had an

intercorrelation of 26, in view of the reliability of the Finger Dexterity quality ratings and a restricted range of Tweezer quality ratings this is difficult to explain

Quality ratings yielded no significant relationships with length of employment or salary ratio, with the exception of Tweezer Dexterity and the former, whereas 64 percent of those who received above-average quality ratings worked for four months or longer, only 39 percent of those rated below average on quality of Tweezer Dexterity remained on the job that long ($D/\sigma d = 3.6$) On the other hand, both Finger and Tweezer dexterity quality ratings yielded reliable contingency coefficients with foremen's ratings (.50 and .24, with .84 the maximum possible)

Time scores on both tests showed significant differences between less-than-seven-day employees and those who remained on the job for more than a year ($D/\sigma d = 4.3$ and 2.5), with differences approaching significance when the former group were compared with the four-month-to-a-year group Correlations between the two tests and salary ratio were .26 and .32 (other data show that the signs should be negative, being time scores), when the two tests were combined, the validity was .39 All three have some statistical significance The relationships with foremen's ratings were not reliable

A further step consisted of applying the previously established critical scores (1.43) to this new group of workers and to the 84 controls who were not tested There was again no relationship with foremen s ratings Of the group who "passed" both tests when the critical score was applied, only 7 percent were discharged in less than one week, while 57 percent were employed for more than a year, for the no-test group the percentages were 23 and 41, for the group who "failed" one or both tests they were 24 and 28 Appropriate critical ratios were clearly significant Salary ratios were .91, .88, and .73 for the three categories just utilized, with the differences again significant

Finger and Tweezer Dexterity Tests are clearly useful in selecting successful watch assembly workers when criteria such as turnover and output (salary ratio) are used

Electrical assembly workers and one type of packer were tested by the USES Division of Occupational Analysis (.750) *pull-socket assemblers, put-in-coil girls,* and *can packers* The groups were 16, 18, and 43 in number, presumably all women although sex is not specified for two groups The criteria were number of pull sockets assembled per hour, ratio of time consumed to complete a unit of work to standard time set

by time and motion study men for put-in-coil girls, and average number of cans packed per hour Only the Finger Dexterity Test was administered to all three groups, with validities of -09, -25, and 26 Put-in-coil girls also took the Tweezer Dexterity Test, the validity being -57 It is interesting to note that the can packers, for whom the correlations of time scores with the Minnesota Manual Dexterity Test and the USES Pegboard were negative, have a positive correlation with time scores on the tests of wrist-and-finger dexterity This suggests that some types of assembly work tend to retain workers who are fast in gross movements but slow in fine, whereas others retain workers who are dextrous in both types of operations presumably the latter would tend to pay more and to be more selective But the finding may be a reflection of a less selective employment policy rather than of less stringent work requirements, for the numbers are small and may have been employed in only one company, and the spread of scores is much greater for the can packers than for the other assembly workers (sigma equals approximately 30 as opposed to 18 seconds) For one of these assembly jobs the O'Connor dexterity tests do clearly have predictive value, apparently that requiring the finest wrist-and-finger movements, for another, somewhat grosser, assembly job it seems to have none (neither did the Minnesota Manual Dexterity Test), for the third and grossest manual job it has low validity of a negative sort, slow test workers tending to be fast task workers The last two relationships may be the result of the operation of chance factors, but the first is too consistent to be the result of chance

Blum and Candee used the O'Connor Dexterity Tests in their studies (105,106) of *department store packers and wrappers,* described under the Minnesota test, finding zero relationships between these tests and output or supervisors' ratings

Occupational differentiation on the basis of wrist-and-finger dexterity is brought out by MFSRI data presented earlier in Table 15 Office workers, particularly those using machines, tend to make scores approximately one sigma above the average employed adult Men who must use their hands skillfully in certain crafts and professions (manual-training teaching, drafting, and ornamental iron work) stand approximately as high Women who assemble small objects (electric meters and instruments) also excel On the other hand, skilled workers to whom technical information and understanding are more important than manual precision (garage mechanics), and assembly workers and operatives whose operations are gross in nature, score no better than the average worker

It would be highly desirable to compare the means of the USES, Blum, and other more recent and more relevant occupational groups with these, but unfortunately this is made impossible by differences in the scoring methods or by doubt concerning the scoring methods, combined with mean scores which seem quite out of line with MESRI norms (e g Blum's mean Finger-Dexterity time of 417 seconds for successful women watch assemblers compared to the mean of 244 seconds for MESRI adult women) Only one comparison seems clearly legitimate, that between Harris dental students (341) and the MESRI norms This shows that the former group stood at the 84th percentile on Finger Dexterity and at the 89th on Tweezer Dexterity, higher than any of the occupational groups for which norms were obtained by the Minnesota project Such a vindication of the clinical judgment of many users of and writers about a test is, unfortunately, all too rare

Job satisfaction has not been related to the O'Connor dexterity tests in any published studies Presumably the tendency is to focus on the worker's need to make a living and on the employer's desire for efficiency, rather than on the mutual need for emotionally adjusted citizens who find satisfaction in their work

Use of the O'Connor Finger and Tweezer Dexterity Tests in Counseling and Selection The studies which have been made with the O'Connor dexterity tests have, like the original investigation in which they were used, been concerned almost exclusively with their use in the selection of vocational or professional students and employees While data from such sources are not only valuable but essential for tests which are to be used in vocational and educational counseling, they are not sufficient We have repeatedly seen that one must also have information concerning the development and maturation of the aptitude or trait in question, in order to be able to apply the test to adolescents, and that information must be available which in other ways throws light on the nature of the characteristic being measured For the O'Connor dexterity tests fairly adequate data are available to help understand the *nature of the trait* it is distinct from others which we are able to measure, and it plays a part in certain types of vocational activities (summarized below) But little is known specifically about its *development and maturation,* apart from the fact that such aptitudes generally mature earlier than intellectual traits This means that caution is necessary in interpreting the test scores of adolescents, although those of 17 and 18-year-olds can probably be used with some assurance of stability

In general, experience with the tests suggests that wrist-and-finger dexterity is likely to be important during the period of *initial adjustment to fine manual jobs*, and that it is likely to be related to success on the job when people with approximately equal amounts of *technical understanding* or trade knowledge are being compared When the latter vary considerably among applicants or employees, differences in them are likely to outweigh the importance of differences in finger dexterity

Commenting on the earlier studies of tests of manual dexterity, Wittenborn (935) has pointed out that the common failure of such tests to prove valid probably lies in the nature of the criteria that have been employed He states

"Most of the criteria which have been employed in the prediction of mechanical ability have been work samples prepared under unusual competition and other atypical conditions which appear to call for a much higher order of spatial visualizing judgment than manipulative ability, e g , the criteria used in the Minnesota study (of mechanical abilities) The so-called motor aspects of mechanical ability cannot be assumed to be of limited significance simply because their significance has not been rigorously demonstrated by suitable studies If investigators employed such criteria as satisfaction in work, duration of employment in routine operations, speed of work, quality of specific operations, piece work output, breakage, fatigability and other factors . it might well be demonstrated that the motor abilities, particularly manipulative ability could be granted a significant role in guidance and selection procedures "

Although his paper was written after the publication of most of the studies reviewed in this chapter, Wittenborn apparently based his remarks almost entirely on the Minnesota mechanical abilities study, for while the gist of his remarks is true, some studies have been made which conform to his suggestions We have seen that some craftsmen whose work requires manual precision and probably some interest in using one's hands excel in fine manual dexterity (ornamental iron workers, manual-training teachers, draftsmen, and dentists) while others whose work requires trade knowledge and insight but no special manual skill (garage mechanics) do not We have seen that watch assembly workers who stand high in fine manual dexterity tend to keep their jobs longer and to produce more than do those who make lower scores on the O'Connor tests We have seen that those whose fine manual skills impress a psychometrist as above average tend to be rated as better workers by

their foremen But they are not important in gross manual work such as packing Wittenborn's insights were excellent, although the state of research was not as lamentable as he thought it

Although Wittenborn was correct in claiming that (fine) manual dexterity is important in some mechanical occupations, its primary importance lies in certain types of semiskilled jobs The principal reason for the apparent uselessness of tests of manual dexterity in guidance and selection lay, not so much in the criteria taken by themselves, as in the types of jobs which were first studied by means of manual dexterity tests, e g , those in the MESRI norm group Other studies discussed in this section have shown that fine manual dexterity is important in simple manual jobs which require rapid wrist-and-finger movements, e g , power-sewing-machine operation and the assembly of small electrical parts, in more complex assembly work requiring both speed and precision, e g , watch assembly, and in other occupations in which rapid manipulation of small objects such as office machines, cash, and the like are involved, e g , office machine operator, bank teller, and typist

The O'Connor dexterity tests can therefore make a contribution to diagnostic and prognostic work in *high schools* and *colleges,* at least for students in their late teens and above In such work they are helpful with students who are considering entering or preparing for professional, mechanical, or office work in which skill with the hands is important, and with others who may enter types of semiskilled factory work in which speed or precision of wrist-and-finger movements is related to stability of employment, earnings, and, probably, satisfaction

In *guidance centers* the tests are useful for the same purposes, and have additional value in employment counseling when initial adjustments are likely to be important Steel and others demonstrated this with their electrical worksample, and Blum with the watch assemblers who remained on the job for less than a week

In *business and industry,* the Finger and Tweezer Dexterity Tests are most useful in the selection of persons who will adapt themselves most readily to speedy or precise semiskilled work They have little to contribute to the selection of skilled, clerical, and professional workers, as those who have completed appropriate training and chosen to continue in the field are likely to be above the critical minimum needed in such occupations

Since there are two O'Connor wrist-and-finger dexterity tests it is in order to ask, finally, whether both should be used or only one will do,

and if so, which one In *heterogeneous groups*, and when rough screening is the objective, one of the tests suffices because of the substantial correlation in such groups Normally the Finger Dexterity Test is to be recommended as a measure of a more commonly used degree of dexterity, but in some situations the Tweezer Test will be more appropriate The Finger Test also has the advantage of having been more thoroughly studied *In homogeneous groups*, and when more refined judgments need to be made concerning manual skill, both tests should normally be used, although local norms and validities will sometimes make possible the omission of one test In any case, it will generally be wise also to use a test of gross manual dexterity, such as the Minnesota, in testing for counseling, in selection testing both should be used in the research stages, dropping the test or tests which prove not to have predictive value in the local situation

The Purdue Pegboard (Science Research Associates, 1948)

The Purdue Pegboard was developed by the Purdue Research Foundation, Purdue University, and published in 1948 as a test of two types of manual dexterity arm-and-hand dexterity of a finer type than the Minnesota Test, and finger dexterity manifested in a more realistic way than in the O'Connor tests Although still new and relatively little studied, both motion study of the test and preliminary data suggest that it merits detailed consideration As pointed out early in this chapter, it appears to tap ability to perform global movements and to eliminate non-essential operations to a degree greater than other manual dexterity tests

Applicability The Pegboard was designed as a group test for and standardized upon adult industrial workers It has since been standardized upon veterans counseled in guidance centers and upon college students, but, like other manual dexterity tests, its development through adolescence to adulthood has not been studied As dexterities generally mature early, it is probably safe to use the adult norms with older high school boys and girls

Content The Purdue Pegboard consists of a 12×18 inch rectangular board with four shallow cups of trays at one end, and two rows of 1/8 inch holes perpendicularly down the middle Fifty easily fitting metal pins are provided, together with 20 metal collars and 40 metal washers made to fit the pins

Administration and Scoring The test is administered with the subject seated at a 30-inch table on which the board is placed with the cups away

from the subject If the psychometrist sits opposite the subject, he must be careful not to let his own hands get near enough to the cups to seem to interfere with the testing The first part tests the right hand, putting the pins in the holes one at a time, the second repeats with the left hand, the third tests both hand simultaneously, the fourth score consists of the first three combined, and a final sequence consists of assembling pin, washer, collar, and washer using right, left, right, and left hands Thus dexterity is tested for each arm and hand, with fingers playing a simple grasping role, ability to perform the same operation with both hands simultaneously is measured, and ability to perform different operations in a co-ordinated way with the two hands simultaneously is assessed As Cohen and Strauss (162) point out, if the worker can effectively merge the two sets of operations in a task such as the assembly test he saves time in the total task, if he must work first with one hand and then with the other, he adds to the time required The assembly test also seems to require finer finger movements than the other parts, which appear to resemble the O'Connor tests The score is the number of pins placed in 30 seconds (sequences 1 to 3) and the number of assemblies made in 60 seconds

Norms The revised one-trial norms (1948 Manual) are for 4138 women applicants for factory employment, 392 college women, 2139 college men and veterans, and 865 male industrial applicants, treated separately, but the numbers are not given in the manual as finally printed Three-trial norms are based on data from 500 college students which made possible the extrapolation of norms for all groups Analysis of data for 900 subjects by previous employment, regional origin, and race failed to reveal any group differences But norms for veterans published by Long and Hill (479) tend to be somewhat lower, particularly the total scores Although these norms are helpful for general interpretation, they throw no light on the vocational significance of the test scores Occupational, especially semiskilled, norms are badly needed

Standardization The test authors stated in the original manual that considerable data had been gathered concerning the test's validity, but that government (wartime) regulations made impossible their publication They added that comparable studies were being made elsewhere, results of which were made available in the revised manual and are described below, under validity

Nothing is said, in the manual, concerning the process of developing the test Reliability data are given .71 for the total score of the combined pin placing tests, and .68 for the assembly test, one trial each (N = 175 to

434) Three-trial reliabilities are estimated to be 88 and 86 To this writer these data suggest that the board should be modified to provide three rows of holes at each side of the board, more pins, washers, and collars, and 90 seconds of working time for each of the pin-placing tests rather than 30 seconds This would not unduly lengthen the test and would give it a reliability more in line with modern standards

Reliability As indicated above, the reliability of the standard one-trial test leaves something to be desired Surgent (807) has confirmed the test author's data with a group of 233 women factory workers

Validity The test being quite new, only one field validation study has as yet been published (807) There will undoubtedly be a number before this book has been long off press

TABLE 16

RESULTS OF VALIDITY STUDIES WITH THE PURDUE PEGBOARD

Test	*No of Trials*	*Job*	*Criterion*	*N*	*r*
Right Hand	1	Light machine operation	Make-up pay while learning	17	56
Left Hand	1	" " "	" " " "	17	23
Both Hands	1	" " "	" " " "	17	21
R + L + B	1	" " "	" " " "	16	31
Assembly	1	" " "	" " " "	16	38
Right Hand	1	" " "	Earnings after learning	17	52
Left Hand	1	" " "	" " "	17	20
Both Hands	1	" " "	" " "	17	07
R + L + B	1	" " "	" " "	16	33
Assembly	1	" " "	" " "	16	38
Assembly	1	Textile quilling	Production Index	28	15
Right Hand	1	Simple assembly, small parts	" "	15	76
Assembly	1	" " " "	" "	15	76
Assembly	3	Radio tube mounters (807)	Pooled overall ratings	233	64

Table 16 gives the results of the validity studies reported in the manual, by permission of Science Research Associates It should be noted that numbers were very small (and the r's therefore not very reliable) in all groups except the last For this group of 233 radio tube mounters, with ratings as a criterion, the validity of the three-trial assembly test was 64 The trend of the other correlations is encouraging but more adequate data are clearly needed

It should be noted that the suggestions for interpretation on the Score Sheet provided by the publisher include artists, chauffeurs, mechanics, musicians, pilots, and others, as well as assembly workers, as groups for which the test should prove useful But no data support these claims, and while pilots, at least, might conceivably make high assembly (co-or-

dination) scores, there definitely is no relationship between manual dexterity and success in flying (214)

Use of the Purdue Pegboard in Counseling and Selection Until further validation data are provided, there is only one kind of situation in which this test can now be used for *counseling* that in which the counselor, or a psychometrist who writes detailed test reports, has a first-hand knowledge of factory jobs acquired by job-analysis experience Such a user of the test may obtain from it clinical insights into the manual dexterities of his clients, which he then subjectively translates into occupational terms Unless this translation is based on intensive job-analysis information it is likely to be dangerously misleading The observer will want to look for efficient use of hands, particularly for global co ordinated movements in the assembly test The nature of the test is such that this writer is confident that specific occupational norms and good validity data can be made available in due course

In *selection,* the test may similarly be used in situations in which decisions have to be made before validation and local norming can be completed Again, job analysis data are needed On the other hand, it is possible and wise, more frequently than most so-called practical men admit, to make immediate decisions on other bases, and to use tests at first only to gather research data which will provide a better basis for similar decisions as the need recurs in the future If research data are not gathered the first time, but the tests are put to intuitive use, then judgmental errors, comparable to those which the tests were adopted to do away with, are perpetuated One type of intuition replaces another

The Purdue Pegboard, modified as suggested in the discussion of reliability, seems to be an extremely promising test for assembly, packing, machine-operation, and other fairly precise manual jobs The analysis of manual work by Cohen and Strauss, discussed in the opening section of this chapter, and the nature and validity of other manual and finger dexterity tests, suggest this It should be valid for a greater and manually more demanding variety of jobs than the Minnesota Rate of Manipulation Test, and should have higher validities than the O'Connor dexterity tests for jobs such as those for which these have proved valid But evidence should be assembled and published

CHAPTER X

MECHANICAL APTITUDE

Nature and Role

THE TITLE of this chapter, and indeed the writing of a separate chapter on this subject, are a concession to practical considerations and to popular usage, rather than an organization of materials dictated by the nature of aptitudes Counselors, personnel men, and vocational psychologists have long been accustomed to thinking in terms of mechanical aptitude They have not defined the term in any strict sense, but have used it operationally to refer to the characteristic or set of characteristics which tends to make for success in mechanical work Tests have been developed which have proved to be reasonably valid for various types of mechanical occupations In one sense, then, there has been some justification for using the term mechanical aptitude But while these practical developments were taking place psychologists were also studying mechanical aptitude in order to ascertain whether it was in fact one trait or aptitude in the limited sense of the term, or whether it was really a combination of aptitudes

The first significant attempts to study, rather than simply measure, mechanical aptitude were carried out by Cox (175) in England and by Paterson and associates (588) at the University of Minnesota Using especially constructed mechanical apparatus which did not lend itself well to scoring, Cox applied factor analysis to his data according to Spearman s two-factor method He isolated a factor which seemed to be of special importance in the mechanical tasks, and therefore might be called "mechanical aptitude", but it was an eductive factor of the spatial relations type, rather than something peculiarly mechanical which might be called "mechanical comprehension"

At about the same time Paterson and his colleagues were carrying out the Minnesota Mechanical Abilities Project, in which they first tried out a number of existing tests, then revised and selected from these to make a definitive study of mechanical aptitude in junior high school boys

As Harvey (950) points out, the Minnesota project was superior in test ideas and construction to the Cox, but was somewhat weaker in theory, for Cox utilized factor analysis theories and procedures which were not yet in use by American psychologists He consequently had not only superior statistical methods but also somewhat more clear-cut hypotheses to guide him in planning his project In the Minnesota project the Minnesota Mechanical Assembly, Spatial Relations, and Paper Form Board Tests were administered, together with the Otis, an interest inventory, and the Stenquist Mechanical Aptitude or Picture Tests (resembling the O Rourke) Data on cultural status, recreational interests, mechanical operations or activities around the home, father's mechanical operations, tools owned by the subject and by his father, mechanical ability required in the father's occupation, and similar factors were obtained The subjects were 150 junior high school boys in Minneapolis Validity aspects of the study will be considered in connection with specific tests, at this point our interest is in the nature of the factors measured by the tests which were selected to appraise mechanical aptitude

Information on this subject comes from studies by Harrell (936) and by Wittenborn (935) Harrell applied Thurstone's centroid method of factor analysis to the Minnesota battery, which he had administered to 91 cotton-mill machine fixers together with more than 30 other tests Five factors emerged, of which two, perception of detail and visualization of space relations, were important in the Minnesota tests The former was demonstrated by repetitions of the tests to be a routine type of ability, whereas the latter played a part only in the earlier administrations of the test to a given subject, Harrell therefore described the spatial factor as the equivalent of mechanical ingenuity Wittenborn applied the same factorial method to the data of the original study In this case the intercorrelations between the Minnesota Mechanical Assembly Test (described later and cited here as the prototype of "mechanical aptitude" tests) and the Minnesota Spatial Relations Test and Paper Form Board were respectively 56 and 49 This suggests that spatial visualization plays an important part in "mechanical aptitude," but does not explain entirely performance on such a test Wittenborn isolated four factors, of which only one, spatial visualization, played an important part in the Mechanical Assembly Test Spatial visualization accounted for 37 percent of the variance in the Assembly Test, this is to be compared with 55 percent of the variance in the Spatial Relations Test, 49 in the Paper

Form Board, and 56 of ratings of the quality of shop work, showing in another way that spatial visualization is important but still only one of the factors which play a part in such instruments as the Minnesota Mechanical Assembly Test

Neither Harrell's nor Wittenborn's studies raise the question, or throw any light on the nature, of the factor or factors which account for the remaining 63 percent of the variance of the Minnesota Assembly Test Neither does another analysis of Cox's mechanical assembly tests by Slater (720), although the last-named investigator agreed with the others in finding no special mechanical factor over and above general intelligence and spatial visualization But this inability to isolate any other factors is in part a function of the types and varieties of tests which are used in the factor analysis one can locate only the factors which are important in several of the tests, and if a factor is important in only one or two tests it may not emerge as significant

Bingham (94 Ch 11) suggests that factors in mechanical success are mechanical aptitude, measured by tests such as the Minnesota Assembly and Spatial Relations Tests, manual dexterity (demonstrated to be unimportant), perceptual acuity (confirmed), and mechanical information The Minnesota study included a measure of mechanical information ("the shop operations information criterion") which had a correlation of 35 with the Assembly Test, but this item was omitted in Wittenborn's analysis, and nothing comparable to it was included in Harrell's data Both authors included the Stenquist Picture Tests which are generally thought to measure mechanical information and which correlate 40 and 46 with the Minnesota Mechanical Assembly Test Although only 22 and 18 percent of the variance in the Stenquist tests is accounted for by the spatial factor (935), and perceptual speed and accuracy plays some part in them (336), they are virtually unanalyzed by Wittenborn and Harrell's studies

Guilford's analysis of a greater variety of tests tried out in the Army Air Forces' Aviation Psychology Program (316,317) provides the answer to the question of what other factors play a part in tests of mechanical aptitude In this analysis, thanks to the inclusion of a test of mechanical information, an aptitude test patterned after the Bennett Mechanical Comprehension Test (described below) was found to be heavily saturated with two factors spatial visualization and mechanical information

What has commonly been thought of as mechanical aptitude, what vocational psychologists have for twenty years known to be partly spatial

visualization, and what some authorities (94) erroneously thought to be also partly manual dexterity, finally emerges in Harrell's and Guilford's studies as a composite of spatial visualization, perceptual speed and acuity, and mechanical information As in the case of Binet's global approach to the problem of measuring intelligence, this lumping together of several aptitudes in one test has had its advantages, for in days when factor analysis was in its infancy reliable and valid tests were developed, effective even though impure, for the prediction of success in mechanical activities With the information and techniques now available purer tests can be developed which will result in a better understanding of both aptitudes and activities, and which will be more versatile in their applicability There is room for doubt as to whether they will be more valid for all purposes, even when combined in batteries, because of the advantages of face validity and the inclusion of specific factors which characterize factorially impure tests depending heavily on job analysis and job content for their items In the meantime multi-factorial tests of so-called mechanical aptitude or comprehension are among the most valid tests available For this reason they are dealt with as such in this chapter, and purer tests of spatial visualization are treated separately in the next, just as tests of manual dexterity were taken up in the preceding chapter

Specific Tests

One of the earliest tests of mechanical aptitude was the Stenquist Mechanical Assembly Test (755), consisting of a long narrow box, each compartment of which contained a mechanical contrivance to be assembled by the examinee The ten items consisted of a mouse trap, a push button, and similar everyday objects Stenquist also developed two picture tests designed to measure the same type of aptitude, but, since manipulation and trial of the parts is impossible in a printed test, it has generally been thought of as being more heavily saturated with information than the apparatus tests As a result of work with Army trade and mechanical aptitude tests during World War I, O'Rourke (277 265) developed a graphic and verbal test of the same type Paterson and associates (588) modified and lengthened Stenquist's Assembly Test as the Minnesota Mechanical Assembly Test for their intensive study of the nature and measurement of mechanical aptitude More recently, Bennett (68) developed his Test of Mechanical Comprehension in order to tap a higher level of mechanical aptitude than the Stenquist, O'Rourke, and

other paper-and-pencil tests already available A totally different type of composite test was constructed by MacQuarrie (504), who combined subtests of spatial visualization and manual dexterity in a test of so-called mechanical aptitude

Of these and other tests like them, the *Minnesota Mechanical Assembly Test,* the *O'Rourke Mechanical Aptitude Test,* the *Bennett Mechanical Comprehension Test,* and the *MacQuarrie Test of Mechanical Ability* have been selected for detailed treatment The assembly test has been chosen as the most adequate of its type and because of the insights which studies using it give into the nature and organization of aptitude for mechanical work, even though it is no longer widely used The O'Rourke has been as thoroughly studied as the Stenquist and other picture tests, and has the advantage of more recent and more extensive norms than most, it is still widely used, although there is room for a well-constructed and up-to-date test of the same type The Bennett is one of the newest but most thoroughly studied and widely used graphic tests of mechanical aptitude, and taps a higher level of aptitude than the other mechanical aptitude tests And the MacQuarrie is not only unique as to content, but widely used and studied, although it could just as well be dealt with under tests of manual dexterity or spatial visualization, it is included in this chapter as a composite test of mechanical aptitude The *Purdue Mechanical Adaptability Test* is also treated, more briefly, as a new instrument of some promise

The Minnesota Mechanical Assembly Test (Marietta Apparatus Co, 1930)

This test was developed as a part of the University of Minnesota's study of mechanical aptitude, in the preliminary work of which it was found that Stenquist's ten-item test had a reliability of only 72 Three boxes, or a total of 36 mechanical items, were used with a resulting reliability of 90 Three of these items have since been omitted, making a total of 33 The Stenquist test having been one of the first fairly good tests of mechanical aptitude, and the Minnesota being a demonstrated improvement upon that, the latter came rapidly into widespread use in clinics and guidance bureaus doing individual testing with adolescent boys, it has not been so extensively used in other situations, because of administration time, wear and tear, and the effects of experience

Applicability Like the Stenquist, the Minnesota Mechanical Assembly Test was designed for use with junior high school boys, and particularly

for the prediction of success in shop courses It was recognized that experience or familiarity with mechanical objects might well play an important part in scores on such a test, even at this age, the Minnesota study therefore analyzed the relationship between a number of environmental factors which reflect or constitute differences in experience, either direct or vicarious, with mechanical objects and processes Two experience items showed positive correlations with the assembly test recreational interests (23) and mechanical household tasks such as electrical repairs performed by the boy (40), on the other hand, ratings of the mechanical ability required by the father's occupation, the tools owned by the boy, and the tools owned by the father, had no relationship with the assembly test scores of the 150 boys of the study (r's = — 11, 14, and 03) Two other relationships are of interest here, one being that with age which is understandably negligible (13) in a group as relatively homogeneous as 7th and 8th grade boys, and the other that with scores on a test of shop information which is moderately high (35) It is noteworthy that the three experience items with which substantial correlations were found probably involve both cause and effect boys with more mechanical aptitude could be expected to choose mechanical hobbies, seek to do household repairs and learn a good deal about shop processes, at the same time, boys who have such hobbies, perform such chores, and learn well in shop courses could be expected to acquire the knowledge to do better than others on a test of mechanical assembly On the other hand, the items which are more strictly environmental, i e , not within the control of the boys but affecting them nonetheless, show negligible relationships with assembly test scores, boys do not choose their fathers' occupations nor decide how many tools their fathers will have, and economic factors and parental ideas probably determine the boys' own tools more than do their desires, but one would expect mechanically inclined fathers who have and use their own tools to have some effect on the mechanical information possessed by boys in their early teens More important, perhaps, than mere possession of mechanical tools and hobbies by the father may be the extent of identification of the son with the father and of father acceptance of the son If this is so, the continua are not experience vs no-experience, but mechanical-father-identification, and non-mechanical-father-rejection, each of which must be combined with son-acceptance and son-rejection in order to describe the emotional as well as material environment which shapes the boy's interests and information Unfortunately, no such refined studies have as

yet been attempted That the mechanical activities of the fathers do not affect the sons seems to indicate that at this age the Minnesota Mechanical Assembly Test is more a measure of differences in mechanical insight (spatial visualization) than of mechanical information

Perhaps this is why Wittenborn's factor analysis (935), cited in the opening section of this chapter, and using the original Minnesota data, failed to isolate any other important factors in this test Harrell (334) reported a correlation of -22 between inexperience and assembly test scores in his study of mechanical aptitudes in adult cotton-mill machine fixers Adults who had had mechanical experience did better than those who lacked it (Harrell also showed that practice on the assembly test reduced it to a measure of perceptual speed and accuracy) We have already seen that Guilford (316,317) found an experience factor in another mechanical comprehension test used with aviation cadets These data lead to the conclusion that in early adolescence tests such as the Minnesota Mechanical Assembly Test are primarily measures of mechanical comprehension (spatial visualization), whereas in late adolescence and adulthood they also tap mechanical information (experience)

Clinical experience with the assembly test has led to the generally accepted conclusion that it is unsuitable and too easy for use with older adolescents and adult men, and too difficult for most women The first is perhaps verified by the AAF study (316) cited above, but none of them have actually been objectively confirmed with the assembly test itself except through data on age differences in the reliability of the test (see below) The most objective evidence, apart from Harrell's data, lies in the norms for various age and occupational groups, which show increasingly higher scores from age 12 to age 19 (raw scores of 232 to 299, the former median being at approximately the 10th percentile for 19-year-olds) But the available data do not tell us whether these increases with age are the result of maturation of spatial visualization or of increased familiarity with mechanical objects As manual training teachers and ornamental iron-workers in the Minnesota Employment Stabilization Research Institute fell midway between the average 18- and 19-year-old boy in the original norms, auto mechanics were slightly lower, and the average employed adult was little more than midway between the average 17 and 18-year-old, the implication is that either the sampling in the adolescent group was skewed toward the upper limits or maturation of spatial visualization plays a greater part in assembly test scores than experience in mechanical activities If this were not so miscellaneous

boys would not surpass skilled mechanical workers It seems more probable that the adolescent sample is not adequate at the upper limits (due to elimination in high school) and that skilled workers surpass 19-year-olds about as much as they do 17-year-olds, that is, by more than one sigma Lacking adequate objective evidence concerning the effects of experience on the assembly test scores of adults it seems wise for practical purposes to agree with Bingham (94 308) and with Paterson, Schneidler and Williamson (590 222) that the varied amounts of mechanical experience which characterize adults make it unwise to use this test with that age group, the theoretical question remains open until better evidence is accumulated

Content The Minnesota Mechanical Assembly Test consists of three boxes containing 33 mechanical objects such as an expansion nut, a hose-pinch clamp, a wooden clothespin a push-button door bell, a spark plug, an inside caliper, and a petcock

Administration and Scoring A fixed amount of time is allowed for work on each object, these being presented unassembled in their compartments Scoring is on the basis of proportion of possible connections made in the allotted time The psychometrist needs to be thoroughly familiar with the assembly and disassembly of the objects, both from studying the directions and from actually practicing with the materials, especially the latter He must know not only how to put the parts together, but what condition they should be in when new for many boxes actually in use contain bent or broken parts and non-standard replacements which change the nature of the task In fact, one problem brought out by World War II testing operations, and not adequately realized when investigations such as the University of Minnesota's study of 150 boys was planned, is the drastic effect on apparatus tests of the wear and tear of large-scale testing In the Air Force program, for example, it was found necessary to assign an officer and several enlisted men to an apparatus control unit at each testing center, their sole function being to make statistical studies of the effects of differences in supposedly identical pieces of apparatus on test scores and to establish correction formulas for raw scores on each apparatus Most of these differences were due to wear and tear through use, as many as 100 men per day being tested by a given piece of equipment

Norms Norms for boys aged 11 to 21 were published by Paterson and associates (588) as a result of the Minnesota Mechanical Abilities Project, and for general adults and specific occupations by Green and others

(306) after the test was used in the Minnesota Employment Stabilization Research Institute Paterson does not make clear the number of cases used in the original norms, which included at least 150 boys in 7th and 8th grades, but unknown numbers at the higher levels Since the test is most useful at the junior high school level this is not a serious limitation The adult norms are based on the Minnesota standard sample of 500 employed adults, the specific occupational groups are small, ranging from 18 draftsmen to 169 manual-training teachers In view of the presumed effects of experience and the suitability of other tests for adult use, the adult norms are of questionable value, they do show the expected group differences, as will be seen below, but these are not as great as one would expect in a good aptitude test, perhaps because of the leveling effects of experience and information with items such as these

Standardization and Initial Validation As has already been indicated, the Minnesota Mechanical Assembly Test was developed as a more reliable edition of Stenquist's test As a part of the intensive study of mechanical abilities carried out by Paterson and associates (588) it was correlated with a variety of other tests and with a number of experience variables in order to throw light on its nature and validity Some of these have already been discussed, in connection with the question of the applicability of the test, others remain to be considered

In the relatively restricted age, but somewhat greater intellectual, range of the 7th and 8th grades the correlation between assembly test scores and Otis I Q was 06 Spatial visualization as measured by the Minnesota Spatial Relations and Paper Form Board Tests, on the other hand, had correlations of 56 and 49 respectively, showing the important role of the spatial factor in mechanical assembly work at this age Correlations with the Stenquist Picture Tests were 46 and 40, as might be anticipated with paper-and-pencil tests of mechanical comprehension

There was no relationship between assembly test score and average academic grades ($r = 13$), but the correlation with ratings of the quality of shop operations was 55, and that with a test of shop information was 35 The higher correlation with operations, as opposed to information, suggests that the test was accomplishing its objective of measuring aptitudes for mechanical work Certainly it predicted success in that much better than in academic work

Two other relationships are of interest, one a correlation of 02 with preference for mechanical occupations, the other a correlation of 42 with scores on a mechanical interest inventory The discrepancy suggests

that the expressed occupational preferences of junior high school boys may not be valid indicators of interest, whereas inventory scores may be, conclusion confirmed by other studies reviewed in the chapter on interests In view of the confirmation of this deduction, it may be concluded that mechanical interests and mechanical aptitude tend to be associated, although the relationship is far from perfect Perhaps the relationship is due to the role of interest in the acquisition of information, and the role of information in so-called mechanical aptitude

Reliability In the original study of the Minnesota Mechanical Assembly Test its reliability was found to be 94 when computed by the odd-even method and corrected by the Spearman Brown formula, based upon 217 junior high school boys (588) In the MESRI project the corrected odd-even reliability was only 79 for 414 adult men, and 68 for 127 adult women (187), the difference presumably being due to the effects of experience in adolescence and adulthood Brush (122) found a corrected reliability of 65 with engineering freshmen For this reason only extremely high and extremely low scores are likely to have any significance for adults In another study using deaf children as subjects, Stanton (719) found retest reliabilities of 74 for boys (N = 57) and 60 for girls (N = 36) after a period of two years, in view of the probability of experience with mechanical objects at that age, and of the known effects of maturation on spatial visualization, these may be taken as not out of line with the other report based on children

Validity The Minnesota Mechanical Assembly Test was correlated with *intelligence* tests in the MESRI project (306), where with adult subjects and the Pressey Classification and Verification Tests the coefficients ranged from 10 to 26, and by Super in an unpublished study with the Otis and NYA youth, in which the correlation was 24 While these coefficients are slightly higher than those reported in the original work with the test, they are low enough to be negligible

No published data on the correlation between widely used *manual dexterity* tests and assembly test scores have been located, but several less used tests in the Minnesota battery yielded low or negligible correlations In his factor analysis of these data, Wittenborn (935) found that manual dexterity did not have an appreciable loading in the assembly test, and Harrell (336), using the same tests and new subjects, confirmed the absence of a manual dexterity factor in this test In an unpublished study of 50 junior high school boys the writer found a correlation of only 05 between assembly test and Minnesota Placing Test scores Apparently

manual dexterity is subordinate to other factors in the task of assembling mechanical objects such as those in the Minnesota test

Although the assembly test was correlated with the Stenquist Picture Tests in the original study, it has apparently not been related to other tests of *mechanical comprehension,* except in an unpublished study by the writer, in which it and the O'Rourke Mechanical Aptitude Test were administered to fifty junior high school boys with a resulting correlation of 65 This is higher than that of 46 reported with the Stenquist in the original study and confirmed by Harrell (334) with adults, although the writer used a similar group of subjects, it is not, however, contrary to what one might expect in apparatus and paper-and-pencil tests designed to measure the same type of aptitude Perhaps it indicates that the O'Rourke more closely approximates a graphic version of the Minnesota than does the Stenquist

The important role of *spatial visualization* in mechanical assembly tests seems to have also been accepted virtually unchecked as a result of the Minnesota project, in which the correlation was 56 between assembly test and Minnesota Spatial Relations Test In the unpublished study of junior high school boys referred to above the writer found the expected correlation of 48 between the assembly test and the Revised Minnesota Paper Form Board, but one of only 25 with the Minnesota Spatial Relations Test In view of the other data, this may be only a chance lack of relationship, which might prove to be higher in other similar samples of the same population Harrell (334) reported a somewhat higher correlation of 35 between Minnesota assembly and spatial relations tests administered to adult factory workers However, the results of his factor analysis agreed with Wittenborn's (935) in describing spatial visualization as the principal factor in the assembly test, and Tredick (869) found its highest correlation among Thurstone's PMA Tests to be with the spatial factor (34, Reasoning was 30, Induction 26, Perception 23)

The correlation between mechanical assembly test and mechanical *interest* inventory scores, reported as 42 for junior high school boys by the original study, was found to be only .10 when the same tests (Minnesota Assembly and Minnesota Interest Analysis) were used with adults by Harrell (334) Whether this is a result of the effects of experience on the test scores, giving them different meaning for adults, or a direct contradiction of the Minnesota findings is not shown by the data, it seems likely that it is to be explained by age differences in experience and its effects on assembly scores

Grades have been used as a criterion by Stanton (749), Tredick (869) and Brush (122) Stanton administered Minnesota Battery A (Assembly, Spatial Relations, and Paper Form Board) to 121 deaf boys aged 12 to 14 The battery validity was 50, and that for the assembly test was 12, using *amount of time spent in shop work* as a criterion This finding is not as favorable as the original 55 reported by the test authors, and the shrinkage seems greater than that normally found between first and subsequent validities, but this may be due to the substitution of a time for a quality criterion

Tredick's study involved 113 freshmen students of home economics at Pennsylvania State College She used the Minnesota Mechanical Assembly Test together with an extensive battery of other tests Her criteria were *semester-point average* and grades in first semester courses in art, chemistry, and English composition The correlations were respectively 11, 17, 16, and — 01, none of which are high enough to be of value

Brush administered the assembly test to 104 freshmen *engineering students* at the University of Maine, correlating results with grades for the first year and for all four years The two coefficients were 28 and 27, both of them reliable Apparently the test has sufficient value for the prediction of success in engineering training to justify its inclusion in a battery, despite the effects of experience by the end of high school In view, however, of the cumbersomeness of administration and scoring, and of the high correlations with paper and-pencil tests of the same type, it is doubtful whether the increased predictive value of a well-selected battery would warrant the time and trouble to include it

Success on the job has not been used as a criterion with the Minnesota assembly test, judging by the lack of such reports in the journals In view of its greater suitability for use with junior high school students than with adults this is perhaps not surprising, it is to be regretted, however, that no follow-ups have been made, to ascertain the relationship between assembly test scores in junior high school and choice of and success in subsequent mechanical employment

Differentiation of occupational groups by the Minnesota Mechanical Assembly Test was demonstrated at the Employment Stabilization Research Institute (223), where machinists scored at the 80th percentile, manual training teachers, ornamental ironworkers, and garage mechanics at the 68th, and draftsmen at the 65th percentiles Workers in less mechanical occupations such as office clerks, machine operators, retail salesmen, and policemen, generally make scores less than one sigma

above the mean of the general population These trends are in the expected directions, although, as pointed out earlier, the mean scores of manual-training teachers and certain other mechanically inclined groups are not as much above the mean as one would anticipate, perhaps because universal experience with the items in the test tends to minimize differences in mechanical comprehension among adults

Occupational satisfaction would seem to be logical criterion against which to validate mechanical aptitude, on the hypothesis that those who are relatively lacking in it would find their work uncongenial and perhaps a strain, while those who are relatively high in mechanical comprehension would solve new problems and master new techniques readily and with zest As an aptitude which is also somewhat related to interest this should perhaps be more true of mechanical comprehension than of most purer "aptitudes" Despite these facts, no known studies have correlated scores on the Minnesota Mechanical Assembly Test with job satisfaction

Use of the Minnesota Mechanical Assembly Test in Counseling and Selection The evidence which has been reviewed in the preceding paragraphs is more adequate concerning the standardization and validation of the Minnesota Assembly Test than are comparable data for most tests, the authors having systematically studied it in a variety of respects Unfortunately it has not been so thoroughly studied since that time, despite Wittenborn's and Harrell's factor analyses One reason for this is the cumbersomeness of the test, not only in administration and scoring, but also in maintenance, another is the proved adequacy of paper-and-pencil tests designed to measure the same factors

Despite these defects, the assembly test is useful with early adolescents whose significant experiences with mechanical items such as those in the test are still largely dependent upon aptitude and interest The effects of maturation upon the principal component, spatial visualization, make the use of adult occupational norms impossible with adolescents The leveling effects of experience, suggested by the decreasing reliability coefficients with increasing age, further complicate the picture and render the scores of older adolescents and adults difficult to interpret

Occupational groups distinguished by high scores on this test include machinists, manual-training teachers, ornamental ironworkers, garage mechanics, draftsmen, and presumably other workers in mechanical occupations, job analysis of which suggests a need for ability to visualize space relations and interest in the acquisition of knowledge about the

nature and operation of mechanical contrivances Whether the superior scores made on this test by adult workers in these fields are due more to aptitude than to experience, or vice versa, does not appear to be important when early adolescents are being counseled, for at that stage the test is largely a measure of aptitude which apparently leads to experience, nor is it especially important when selecting adults for a related type of work, for in such a case present ability to do the work is important, regardless of its basis It is only when long-term adjustments and ability to learn are important that it is necessary to distinguish between experience and aptitude as causative factors in assembly test scores

School and college use of the assembly test has proved feasible, the test having predictive value in junior high school, high school, and engineering college courses In view of the equally high validities of other tests, and the little added to battery validity by this test at any save the junior high school level, it is doubtful whether the time and trouble required for its use are justified The test may be of considerable value, however, in the clinical study of the aptitudes and experiences of special cases

Guidance centers and clinics are most likely to find the test valuable in this type of case When a client's experience with mechanical objects is in need of further study because of lack of mechanical outlets, or when his aptitudes as measured by other tests of spatial visualization, mechanical information, and manual dexterities seem out of line with his experience, then administration of the assembly test by a skilled examiner may prove fruitful The ease with which the subject approaches the apparatus, the familiarity displayed by his examining and assembling of them, his confidence in his ability to complete the assemblies in time, his reactions to difficulties and failure, his incidental comments concerning the test and related matters during and after testing, all provide material in addition to the actual score which a skilled psychologist can piece together in order to obtain a truer picture of the client s aptitudes, interests, and experiences

Business and industrial use of the Minnesota assembly test is probably unwise because of its unreliability with adults, the leveling effects of experience, and difficulties in administration It is true that it can have some value in indicating present mechanical skill in job applicants, but if these are in a skilled category trade tests are more appropriate and valid, and if they are semiskilled manual and spatial tests will prove more economical and more valid

In summary, then, the Minnesota Mechanical Assembly Test is im-

portant primarily for historical reasons and for the insight the studies with it give into the nature of mechanical aptitude, its practical use is limited primarily to the clinical study of special cases, especially in adolescence

The O'Rourke Mechanical Aptitude Test, Junior Grade (Psychological Institute, 1926, 1940)

The O'Rourke Mechanical Aptitude Test was developed after World War I, as a result of the test author's experience with the Army Mechanical Aptitude and Army General Trade Tests, incorporating essentially the same items (277 265 ff.) According to Fryer, the original work by Rice was carried further by O'Rourke and Toops in the Army, and the former continued to work with the test during the early 1920's It was subsequently restandardized in three forms with Tennessee Valley Authority workers (612) Unfortunately none of the work done by O'Rourke has been published, leaving us entirely dependent for our understanding of its development upon Fryer's brief account of its origin, Toops' dissertation, O'Rourke's and Pritchett's unpublished dissertations, and the sketchy data published on the test form and scoring key

Applicability The civilian edition of the test prepared by O'Rourke was first used with boys in their late teens who were interested in entering "mechanical" occupations Just which occupations were included under this heading is not indicated but the fact that his contemporary, Thorndike, classified wrestlers as mechanical workers (828 24) suggests that a word of caution in accepting the designation may be warranted The group on whom the military form had been standardized were draftees, therefore mostly young men, the civilian group were aged 15 to 24, were no longer in school, and none of them had completed more than one year of high school The second standardization of the civilian form was on workmen who applied for mechanical jobs with the Tennessee Valley Authority Again the term "mechanical" is not specifically defined, but a list of occupations for which mean scores are provided includes apprentices as well as journeymen, in such fields as automobile mechanics, boilermaking, carpentering, machine-shop, painting, and even textile manufacturing This suggests that O'Rourke's definition of the term mechanical is as broad when applied to occupations as it is when applied to the types of information which make up the content of his test

Most important, from the point of view of the applicability and use of

the test, is the fact that the means and standard deviations for older adolescents without mechanical training (the original norm group), for adult men with mechanical and other skilled and semiskilled training (TVA), and a group of 785 adult men in a WPA educational program in California (329) are approximately the same This suggests that the test is probably equally applicable to older adolescents and to adults In view of the evidence which suggests an effect of experience on the Minnesota Mechanical Assembly Test this seems surprising, but it is perhaps due to the fact that by their middle teens boys who have mechanical aptitudes and interest learn as much about the tools and processes tested as they ever will It is conceivable that the additional trade knowledge gained after that time is in specialized fields and of an advanced type which does not affect general "mechanical" information such as is tapped by this test As age differences have not been studied as such, it is not possible to give an adequate answer to the question of the effect of age and experience on this test

Content The test consists of two parts The first is pictorial, the subject matches pictures in order to show which tools and other objects are used together The second part is verbal, it is a multiple-choice test concerning tools, materials, and processes As stated above, the term "mechanical" is broadly conceived to include mechanics, electricity, carpentry, cabinet-making, painting, printing, surveying, and other activities, the items being of a type which might be learned in everyday activities, without actual technical training No rationale is offered for the proportions allocated to each field, although these vary greatly Table 17 shows, for example, that Form A includes 19 auto mechanics items, 16 carpentry, and 19 electrical, only 1 drafting, 1 brick laying, and 1 painting, but no plastering or shoe repairing items At the same time, Form B contains 24, 16, and 9, 4, 0, and 0, and 1 and 1 items in each of these same categories This seems likely to lessen the equivalence of the three forms, although no notice seems to have been taken of the fact

Administration and Scoring The two parts require 30 and 25 minutes of working time, respectively, with a brief practice period at the beginning Both parts must be used, no norms being available for the subtests The test requires somewhat more supervision than the average group test, because it is arranged in folder form which confuses many examinees, and because the time limits are excessive for many high school students who finish Part I and proceed to work on Part II before instructed to do

TABLE 17

O'ROURKE MECHANICAL APTITUDE TEST

Number of Items in Each Form of the Test by Different Occupational Activities

	Occupational Activity	Form A			Form B			Form C		
		Part I	Part II	Total	Part I	Part II	Total	Part I	Part II	Total
1	Auto Mechanics	10	9	19	13	11	24	10	15	25
2	Carpentry	9	7	16	10	6	16	6	7	13
3	Electrical	7	12	19	1	8	9	2	8	10
4	Mechanical	4	9	13	5	3	8	5	5	10
5	Plumbing	5	7	12	2	5	7	5	3	8
6	Machinist	2	4	6	4	5	9	1	–	1
7	Mechanical Comprehension	–	2	2	1	4	5	2	4	6
8	Drafting	1	–	1	1	3	4	3	2	5
9	Metal Working	1	–	1	2	3	5	–	3	3
10	Cabinet-Making	1	2	3	1	2	3	–	2	2
11	Wood-Cutting	–	–	–	2	2	4	1	1	2
12	Forge Work	2	–	2	–	–	–	2	1	3
13	Foundry	–	2	2	–	3	3	–	–	–
14	Surveying	1	1	2	–	2	2	1	–	1
15	Pick and Shovel Work	–	–	–	1	–	1	2	1	3
16	Painting	–	1	1	–	–	–	1	2	3
17	Plastering	–	–	–	–	1	1	1	2	3
18	Welding	–	1	1	–	2	2	–	–	–
19	Bicycle Repairing	1	–	1	–	–	–	–	1	1
20	Glazing	–	–	–	1	–	1	1	–	1
21	Printing	–	2	2	–	–	–	–	–	–
22	Shoemaking	–	–	–	1	–	1	–	1	1
23	Stationary Engines	–	1	1	–	–	–	1	–	1
24	Steel Construction	–	–	–	–	–	–	–	2	2
25	Brick Laying	1	–	1	–	–	–	–	–	–
26	Farming	–	–	–	–	–	–	1	–	1
	Total	45	60	105	45	60	105	45	60	105

so The test is frustrating to girls and to boys without mechanical inclinations who feel that it is unreasonable to require them to sit for an hour over questions they cannot answer in any amount of time Scoring is by means of an old-fashioned stencil which is placed against the answer spaces in the test booklet Unless the test is revised for special answer sheets and punched stencils it is likely to lose its market to some more administrable test of the same type

Norms The original norms, published in 1926, were based on 9000 boys aged 15 to 24 who were "entering mechanical occupations" As has already been pointed out, the meaning of this phrase is not made clear the individuals in question may have been mere applicants, many of

whom were rejected, or they may have been successful trainees, the occupations may have been semiskilled, or they may have required considerable insight and knowledge That their educational level was not high is shown by the fact that none had gone beyond the first year of high school, but in 1926 that meant only that they had as much education as the average adult male

The TVA norms were based on 70,000 men who applied for so-called mechanical jobs, an unusually large standardization group These norms differ little from the earlier set, the mean of the adolescent group being a raw score of 198, that of the adult group 190 (equivalent to the 54th percentile in the early norms) The lowest quartiles are 162 and 137, respectively, and the third quartiles are 245 and 242 The adult group includes more low-scoring cases than the adolescent group, perhaps because of loss of speed with age, perhaps because of regional differences in populations The age distribution of the adult group is not given, but the rural southern localities from which the latter of the two groups came, as described by Pritchett (612), suggest that at least the latter reason may apply Hanman s (929) study was based on California men aged 20 to 65, with a mean age of 40, and found a distribution of scores like that of the original norms, which suggests that age differences are probably not the cause

O'Rourke's manual also provides norms for 33 specific occupations in the TVA population, ranging from auto-mechanic apprentices and journeymen, through foundrymen and plasterers, to textile worker journeymen (just which of the Dictionary of Occupational Titles' more than 1800 different textile workers is not specified), welders and woodworkers, This is an unusually large number and variety of occupations for which to provide norms, and in this respect O'Rourke has set an example for other test authors Unfortunately, however there are serious hidden defects One of these poor and at times even meaningless occupational classification, has just been pointed out it is impossible, without more data on some of the jobs, or without reference to a standard classification system such as the Dictionary of Occupational Titles, to know what the norms mean A second defect is the provision of only means and sigmas, this is much less serious, but if the numbers are adequate, more specific norms could easily have been provided With no indication of the numbers in each category, it is impossible for the test user to know whether, as in the case of the MESRI occupational norms, the data are merely suggestive, or whether they can really be used for norms The importance

of this point is brought out by the fact that the means are sometimes very close together, and sometimes even in the reverse of the expected order For example, millwrights make a mean of 199 whereas for machinists the mean raw score is 211, and truck and tractor operator apprentices score three points higher than journeymen Differences such as the former probably reflect in part the composition of the test which, as pointed out in the discussion of content, is very unevenly weighted for the various fields it taps, the latter type of difference is presumably due to sampling errors Both lessens one's confidence in the value of the norms, which may be useful as a rough indication of validity for the directional counseling of adolescents (see below under Occupational Differences) but which can hardly be used for the counseling or selection of individual adults without more descriptive data and detail

Standardization and Initial Validation According to Fryer (277), in their work with the Army Mechanical Aptitude and General Trade Tests, and in their subsequent dissertations with these instruments, O'Rourke found correlations between the two Army tests and ratings of the mechanical ability of high school boys of about 30, and O'Rourke and Toops found correlations with school grades which ranged from 16 to 41 Correlations with Army Alpha were 30 for the Army Mechanical Aptitude Test and 42 for the Army General Trade Test, based on a group of 208 8th grade boys Correlations of the same tests with the Stenquist Mechanical Assembly Test were 41 and 33, with the Stenquist Picture Test I 44 and 27, and with the Stenquist Picture Test II 46 and 33, based on 145 8th grade boys

The two Army tests were validated on student-soldiers awaiting return to civilian life after World War I and on junior high school groups studied by O Rourke and Toops, data from whose dissertations are provided by Fryer (277 Ch 8) The former were rated for achievement after the completion of courses, the numbers ranging from 24 to 61 per course For the automotive course the validity coefficient for the Aptitude Test was 05, but in electrical and machine-shop courses they were 50 and 43 Comparable validities for the Trade Test were 20, 53, and 47 For the 208 junior high school boys the two tests had correlations of 33 and 41 with grades, for the 100 boys who subsequently entered high school the validities were 16 and 32 when only the reliable grades were used

Work dealing specifically with the standardization and validation of the final published version of the O'Rourke test has not been published Fryer (277 270) states that the O'Rourke is a modified version of the

Army Mechanical Aptitude and General Trade Tests Part II, for example, consists of 60 multiple-choice questions rather than 50 one-word completion items as in the Trade Test from which it originated In view of these changes, considerable restandardization must have been done. All O'Rourke tells us, however, is that the published form is based on 9000 fifteen- to twenty-four-year-old males no longer in school and entering mechanical occupations There is nothing on reliability Concerning validity, the manual states that "correlations reported between test scores and ratings in vocational courses are as high as 84, between test scores and ratings in school vocational classes 83 " These are, it should be noted, cited as maximum validities obtained, they are considerably higher than the best validities reported for the two Army Tests, they are also considerably higher than the validities of single tests generally prove to be when they are cross-validated They cannot therefore be taken as indices of the actual validity of the O'Rourke test Judgments concerning its validity must be based solely on inferences from the Army tests and on the published reports of subsequent investigators

Reliability The reliability of the O'Rourke Mechanical Aptitude Test has never, so far as this writer has been able to ascertain, actually been established Bingham (94) estimates that the standard error of measurement does not exceed 18 raw score points, or less than one-half sigma, but this is just an estimate As the 40 item Army General Trade Test had a reliability of 98 (277 268) it seems probable that the longer O'Rourke is also reliable

Validity The intercorrelation of Parts I and II of the O'Rourke is 52 (493) The O'Rourke has been correlated with *intelligence* in an unpublished study by the writer, who administered it and the Otis S A Test to 108 high school junior and senior boys, the resulting coefficient being 23 Sartain (669) reported a correlation of 16 for the same tests administered to 46 aircraft factory inspectors

Other mechanical comprehension tests with which the O'Rourke has been correlated include the Stenquist (686), data having been obtained from 114 7th and 8th grade boys $r = .375$, the Minnesota Mechanical Assembly Test (unpublished data of the writer's) administered to 50 7th grade boys $r = .65$, and the Bennett Mechanical Comprehension Test (493), used with 147 high school and defense-training students $r = .55$ Sartain (669) also reported an r of 55 between the O'Rourke and the Bennett McDaniel and Reynolds (493) found a correlation between the O'Rourke and the MacQuarrie Test of Mechanical Aptitude as high

as .51, but in Scudder and Raubenheimer's study (686) of junior high school boys the correlation was only 01, a difference which it is difficult to explain without more data Sartain's (669) data tend more toward a lack of relationship with the MacQuarrie (r = 20)

The only *spatial visualization* test with which the O'Rourke has been correlated is the Revised Minnesota Paper Form Board, in studies by Tuckman (876), Sartain (669), and the writer (unpublished data) These coefficients were 40, 09, and 44, the subjects of the first study being clients of a Jewish Vocational Service, those of the second experienced factory inspectors, those of the last high school boys in the junior and senior classes The differences in degrees of relationship are probably due to differences in mechanical experience

Interests were related to the O'Rourke by Leffel in an unpublished master's thesis (460) The subjects were 121 boys in the junior and senior years of high school The correlations with Strong's Vocational Interest Blank were 42 for the Chemist key, 46 for the Engineer key, 27 for Mathematics and Physical Science Teacher, and approximately − 25 for the keys for Social Science Teacher, Lawyer, and Certified Public Accountant

Grades were used as criteria in a study of 114 7th and 8th grade boys by Scudder and Raubenheimer (686), with a reported correlation of 15 between the O'Rourke test and grades in shop courses McDaniel and Reynolds (493) used instructors' ratings of 149 high school and defense-training-course students The multiple correlation coefficient between the battery and ratings was 47, the validity of the O'Rourke alone was 26, no other test having a closer relationship with the criterion In a third study, Ross (651) tested an unspecified number of machine-tool trainees in the Parker Defense Training Program at Greenville, South Carolina He established critical scores for the tests used, that for the O'Rourke being 175, this score would have eliminated 67 percent of the failing trainees, together with only 7 percent of the successes The criterion of success was grades in the training courses The correlation with scores on the O'Rourke was not ascertained A study conducted in an aviation machinists school by the U S Navy during World War II (785 247) used grades as a criterion The validity was 65 Other Navy studies used custom-built tests of similar type It should be noted that, although the O'Rourke Test of Mechanical Aptitude is thus shown to have some validity for predicting the quality of work done in mechanical courses, and about as much validity as other available tests, it gives a considerably

less accurate estimate of achievement than is suggested by O'Rourke's partial data

Success on the job was studied with aircraft factory inspectors by Sartain (669) and with Tennessee Valley Authority workmen by Pritchett (612) Sartain's report is unfortunately very brief he provided no information as to the type of inspection or materials inspected, although there are probably very important differences in the psychological and technical demands made upon inspectors of fuselages on the one hand and of engines on the other, the sex and ages of the workers are not specified, representativeness of the sample is assumed without evidence other than the facts that "some of them" were relatively new and "many" were among the most experienced in the department Two criteria were used ratings (we are not told of what) in a refresher course (subject-matter not specified), the two instructors of which were in most cases familiar with the job performance of the inspectors, and merit ratings made by supervisors during the year following the refresher training There were 46 employees in the early group, and 20 still on the job one year later The correlation between ratings by the two instructors was .77, which compares favorably with the reliabilities of ratings in general When correlated with the combined merit ratings made during the subsequent year, the coefficient was .42 In one sense this is a reliability coefficient, because both sets of ratings were based partly on job performance, in another sense it is a validation of the ratings given in refresher training, for it shows that they were positively related to ratings of subsequent job performance Sartain did not report the correlations between tests and merit ratings one year later, perhaps because the number of cases was by then reduced to 20 With ratings in refresher training the correlation for the O'Rourke test was .24, as compared with .32 for the Bennett Test of Mechanical Comprehension, .17 for the Minnesota Paper Form Board, and .64 for the Otis

In view of all the unknowns in this investigation, ranging from the nature of the work, through the characteristics rated, to the similarity between refresher training and the job itself, it is difficult to evaluate Sartain's findings It may be safe to assume, in view of the findings of other studies, that the high correlation between intelligence and ratings was due to intellectual factors which are more important in training than on the job, moderately high correlations between spatial relations tests and ratings suggest that the inspection job, or at least the refresher training related to it, required ability to visualize spatial relations, the

lower validity of the O'Rourke suggests that, at this stage of experience, general mechanical interest and information are less important than spatial visualization and intelligence in this type of work Generalization to possible use of the O'Rourke in the selection or guidance of inexperienced workers is impossible, however, not only because the type of inspection work and training was not described, but also because the role of the factors measured by the O'Rourke may be quite different at the novice as contrasted with the journeyman stages This was seen to be the case, for example, with manual dexterity tests and department-store wrappers

Pritchett's dissertation (612) might be expected to deal more directly with job success His data are based on the administration of the O'Rourke to 70,000 applicants for skilled jobs with the TVA The criteria were efficiency ratings, promotions, demotions, and lay-offs But no evidence is given, beyond a brief statement to this effect

Occupational differences in scores on the O Rourke Mechanical Aptitude Test are shown in the manual, by data obtained in administering the test to applicants for TVA employment High-scoring occupations include journeyman electricians, machinists and sheetmetal workers with mean raw scores equal to approximately one standard deviation above the general mean (209 to 228), apprentices in each of these fields are generally somewhat lower than journeymen Low scoring occupations include watchmen, foundrymen textile workers, and plasterers, all more than one sigma below the general mean (raw scores from 140 to 147) It is noteworthy that auto mechanics, mechanic-millwrights, plumbers, and carpenters make mean scores not significantly higher than the general average This is presumably a reflection of the fact that the items in the O'Rourke test sample a variety of skilled trade subjects, some fields being more heavily weighted than others We have seen that mechanical and electrical items are most numerous in Part II, and that foundry and cabinetmaking are barely represented, it is only logical, then, to find carpenters and machinists making higher scores than foundrymen and carpenter-finishers As a trade test for selecting skilled workers the O'Rourke is, therefore, inadequate there are too many irrelevant items for most trades, and not enough relevant for others As a test for measuring underlying aptitude in experienced workers it leaves much to be desired, since an electrician's score, for example, is heavily weighted by his experience with many items, whereas a foundryman's score is relatively little affected by his training and experience, only one item in Part II being directly relevant As a general mechanical aptitude test for

untrained adolescents, on the other hand, the test seems much more appropriate, for this group has more opportunity to follow up interest in and aptitude for mechanics and electricity than for foundry work, and differences in general information in these areas may more legitimately be taken as indicative of differences in aptitude and interest It would seem worth while, however, to develop a mechanical or technical information test which sampled each of the major fields adequately enough to yield part scores which would be diagnostic of special aptitude or proficiency (depending upon the age of the examinee) in the various fields

In an unpublished master's thesis, Leffel (460) classified 121 high school juniors and seniors according to the occupational fields which they named as their objectives The boys who planned to enter technical professions or semi-professions made significantly higher scores on the O'Rourke than did those who planned to enter other fields, while those who planned to enter social science occupations made significantly lower O'Rourke scores

Job satisfaction has not, so far as the writer has been able to determine, been used as criterion for the O'Rourke test It would seem logical to expect those who have a high degree of mechanical aptitude to be dissatisfied without outlets for it, and to expect those whose work requires more such aptitude than they have to be dissatisfied with their too-demanding work situations

Use of the O'Rourke Mechanical Aptitude Test in Counseling and Selection The findings discussed in the preceding sections show that the O'Rourke Mechanical Aptitude Test is only slightly correlated with intelligence, and that it has a moderately high correlation with other mechanical comprehension tests, with tests of spatial visualization, and with measured interest in mechanical and scientific activities It is therefore possible that the acquisition of mechanical information such as is measured by this test is the result of spatial aptitude, technical interest, and presumably opportunity, unfortunately, no studies have been made which prove causation From the practical point of view, however, the relationships between the O'Rourke and tests of these other factors is low enough to warrant using it in a battery of tests for appropriate persons and for suitable purposes

Changes in scores with age after mid-adolescence have not been brought out by the norms, but this may be due to failure to make a refined analysis of age differences, the only data are the similarity of the means of older adolescent and adult groups This seems surprising in an

information test, but it could be due to the fact that the items in the test tap a low level of information which is generally acquired from miscellaneous sources during adolescence, rather than the higher level of technical information which is learned in training or on the job That this is so has not been demonstrated, but the existence of two such levels of information has frequently proved to be a good working hypothesis in test construction, and its use by O'Rourke is implied in the sub-title, "Junior Grade"

The occupational significance of the O'Rourke test can only be broad, because of the unbalanced heterogeneity of the so-called mechanical items it contains and because of the resulting dislocation of the occupational norms Evidence with both adolescents and adults indicates that the test has some value in distinguishing those who have some aptitude for technical work from those who have little such aptitude, it does not, however, make possible differential diagnosis or prediction within the field of technical work

In schools, technical institutes, and colleges, the O'Rourke test should prove most useful with those who have had no training and no systematic experience in technical fields In such instances it will reveal the extent to which the person in question has sought and utilized opportunities for the exercise of technical aptitudes and interests It will not help in determining in which of the various technical fields he is likely to do best or find most satisfaction, but it does have value in general directional guidance It can normally be expected to improve the selection of high school students who will do well in technical courses, but is not likely to predict success as well as the test manual implies

In guidance centers and employment the possible uses of the test are about what they are in educational institutions It can be useful in selecting promising young trainees or entry workers for industrial employment, supplementing the history of mechanical and related interests and activities

In industry the O'Rourke test is also useful in selecting young people for entry jobs and for training opportunities, as a measure of previous exposure to and profit from incidental technical experiences Although it cannot properly be used as a trade test, it has been shown to have some value as a screening device even for experienced workers on technical jobs, when large numbers have to be employed and the evaluation of experience is difficult In any case, the O'Rourke should be supplemented by purer and less easily contaminated tests of aptitudes such as intelli-

gence and spatial visualization, and, in the case of experienced workers, by trade tests and it should go without saying that personal data should also be utilized

The Bennett Test of Mechanical Comprehension (Psychological Corporation, 1940)

This test of mechanical aptitude was developed after a survey of existing tests of mechanical aptitude led the author to the conclusion that there was a need for a test which would measure a higher order of mechanical aptitude than that assessed by available tests The facts concerning the Minnesota and O Rourke tests, summarized and discussed in the preceding sections of this chapter, partially substantiate that conclusion, as would those concerning other tests of mechanical aptitudes were they similarly treated

Applicability There are three forms of Bennett's test AA, designed for high school students, engineering school applicants, and other relatively untrained and inexperienced groups most widely used and therefore selected for detailed treatment in this chapter, BB, more difficult and designed for use with engineering school applicants, candidates for technical courses and applicants for mechanical employment, and W1 (developed in collaboration with Dinah E Fry) designed for use with high school girls and women An attempt was made to devise items appropriate to the aptitude and experience of each of these types of groups In the case of the women's form, for example, items used embody what seem to be the same types of physical principles, but the objects and situations are such as are more commonly encountered by women than those in the men's forms, they involve the kitchen and the sewing room more than the shop and the garage That this goal of devising items suitable to the group in question was reasonably well attained is illustrated by the fact that 9th grade boys make raw scores which range from 5 to 54, with a mean of 31, whereas 12th graders' make scores ranging from 11 at the first percentile to 57 at the 99th, the mean being 39 As the total number of items is 60, this demonstrates that most of the items are actually working at this age range, and that the improvement which takes place with age in adolescence does not make the test too easy Freshmen engineers, on the other hand, make raw scores of 56, 57, and 59 at the 90th, 95th, and 99th percentiles, and a raw score of 47 at the 50th percentile, showing that that test is so easy for freshmen engineers that the most able cannot show the true extent of their ability Form AA

is suitable for engineering school applicants, as the author states, for in such a selection program the principle objective is to screen out those who are too weak rather than to locate those who are unusually able; in a scholarship program, however, it would be better to use Form BB, thus achieving discrimination at the top and locating the most able Another indication of the suitability of the special forms lies in the fact that women's scores average about 12 points lower than the scores of comparable men on the men's form (69)

The question of the effect of having studied physics upon scores on a test such as Bennett's is frequently raised, since the items measure understanding of and ability to apply physical principles Two studies have investigated this problem, both reported in the manual In one study 315 applicants for defense-industry training answered a question concerning previous training in physics The 220 persons who had had such training made a mean score of 41 7, while the 95 reporting no training made a mean score of 39 7, difference which was in the expected direction but not great enough to be *statistically* significant Expressed in percentiles, one group was at the 60th and the other at the 50th percentile, both of which can be thought of as average As four raw-score points (equal to less than one-half sigma) generally make less difference than this in percentiles, the difference can be thought of as *practically* insignificant also A similar analysis was made of data obtained from 1471 candidates for positions as firemen and policemen in New York City, the biserial r between having had training in physics and score on Bennett's tests was 26, and the difference in the means was again four points or less than one-half of one standard deviation

Content The items of the Bennett Test of Mechanical Comprehension, unlike those of the O'Rourke, are objects which are almost universally familiar in American culture airplanes, carts, steps, pulleys, windlasses, see-saws, and cows In this respect the test is presumably less subject to the effects of differences in experience and environment than is the O'Rourke This is probably also true of what the examinee must do with the objects in order to take the test, for the tasks require comprehension of the nature, operation, and effects of various physical principles rather than knowledge of specific tools or items of equipment and their uses To put it concretely, in Bennett's tests it is not a matter of what to use a pulley for, but rather one of how weight is distributed on pulleys when they are used The only knowledge needed for the latter type of item is an idea of the general nature and use of pulleys, the

answer can be found by logical analysis of the problem, that is, by mechanical comprehension There are a total of 60 such items The existence of a sex difference equal to one and one-half sigma (manual) shows that cultural factors affect even this test, but it seems likely that, for a given sex, they are less important, as witness the data on physics training

Administration and Scoring The test has no time limit, being designed as a power rather than as a speed test The majority finish in less than 25 minutes, and a 30-minute time limit is ample for almost any group Booklets are used a number of times, a special answer sheet being provided for responses Sample problems help orient the examinee to the methods and forms Scoring is by means of stencils, either by hand or in the IBM scoring machine Both administration and scoring are simple and expeditious

Norms For Form AA three sets of norms are available, one for educational groups, one for occupational groups, and one for women For Form BB they consist of data for technical educational groups and applicants for mechanical work The women's form has educational and occupational norms

The educational norms (Form AA) are for each of the four years of high school, each year being based on from 300 to 833 boys, for technical-high school seniors, for introductory engineering school freshmen, there being from 402 to 613 cases in each of these last groups The means increase from year to year, and from less-selected to more highly selected groups

The so-called industrial norms are in some cases more truly educational, as when they are based on candidates for WPA mechanical courses or on clients of a veterans guidance center (veterans are not an occupational group, but a cross section of young men) In other cases they are marginally occupational, being based, for example, on candidates for positions as policemen and firemen (occupational norms could have been obtained by excluding those not actually appointed), candidates for apprentice training, candidates for engineering positions (as their average education equalled two years of college they could not be considered engineers without substantial appropriate experience), and applicants for jobs as mechanics' helpers, unskilled laborers, and leadmen Only two groups are truly occupational, the paper-factory workers and bus and street-car operators The numbers in each of these categories range from 145 candidates for engineering positions to 2217 applicants for employ-

ment as mechanic's helper The two strictly industrial or occupational groups number respectively 1637 and 734

While the numbers are generally sufficiently large, there is no way of knowing how representative they are the schools and colleges from which the educational norms were obtained are not specifically identified, although some can be guessed from the list of acknowledgments in the manual, and the pre-occupational and occupational norms are not described as to location, number of companies, age, or other variables, although here again one can identify some groups by deduction and obtain further information from the original studies the defense-course trainees, for example, appear to be Moore's (536) cases, while the paper-factory workers are a group tested in a Savannah, Georgia plant but not described in any detail (72)

Women's norms for Form AA are based on one small group of college freshmen (N = 111), a moderately large group of wartime applicants at an employment agency (N = 238), and 1090 trainees in an airplane factory With no other information concerning the college and employment agency groups their norms are of little value, for other colleges may have different types of students and women job seekers are not the same in peace and in war The airplane factory workers constitute a large and, judging by other information about women workers in wartime airplane factories, heterogeneous enough group so that they can be of some use The limited norms for women are perhaps not too important in any case, since women do not ordinarily compete for mechanically demanding jobs in peacetime and, when they do, must hold their own with men In a period of industrial mobilization for war production the opposite is, of course, true, and an instrument which can select mechanically apt even though inexperienced women is of great value

As the manual has been revised by its author in order to keep it up to date (in less detail than one might wish) and he and his associates have continued to publish new studies involving the test, it can probably be assumed that the defects in the norms will be progressively minimized, and that in due course both more representative samples and more adequate descriptions of the samples will be made available

Standardization and Initial Validation As described in the manual preliminary work with this test consisted of preparing rough sketches of proposed items and trying them out on various types of persons After elimination and revision of items 75 were tried out in booklet form As a readily available criterion for the retention of items in the test, scores

on three existing tests of mechanical aptitude were combined with the Bennett scores these were the MacQuarrie, the Detroit, and the Revised Minnesota Paper Form Board The responses of the highest and lowest scoring 27 percent of a group of 283 applicants for skilled technical training were used for item analysis, just as in an item analysis of one test by itself but with the advantage of additional items designed to measure the same trait to help differentiate the most able from the least able candidates This is therefore neither an internal consistency validation nor a validation against existing tests, but rather a mixture of the two As a result of this procedure the number of items was reduced to 60, plus two easy items which were retained as practice questions having survived such an analysis, these items can be presumed to be measuring the same trait or constellation of traits, to be measuring something resembling what others have called mechanical aptitude, and if these other tests have some validity for measuring promise in mechanical work (as they do), to have some validity as a test of mechanical aptitude Such indirect proof of validity is not satisfactory in and of itself, but it suffices as a first step, the successful taking of which then justifies the labor of validating against occupational criteria

Reliability The only reported reliability coefficient located by the writer is that given in the manual, 81 for a group of 500 9th grade boys, calculated by the split-half method This is sufficiently high, especially for such a homogeneous group, it would presumably be higher if the age and ability range were greater

Validity Because of the strength of its rationale and the consulting activities of its author, the Bennett Test of Mechanical Comprehension has been used in a surprisingly large number of studies, including several in the Army, Navy and Air Force which have not yet been reported in the general literature Criteria used have included not only other tests, but grades and supervisors' ratings, output and other objective vocational criteria have not, however, as yet been utilized as criteria, perhaps partly because the test was designed and used primarily for jobs above the semiskilled level in which success cannot often be judged by production records

Tests of intelligence which have been correlated with Bennett's have been summarized in a table in the manual Of special interest are the correlations of 25 and 45 with the Otis S A Test based on 156 high school and on 292 defense training students The manual does not indicate the age or grade range of the high school students, but the low

correlation may be due to homogeneity, the higher correlation for the defense-industry trainees is presumably due to greater ranges of education and age Other correlations with the Otis test have been reported by Sartain (669 671), who found a relationship of only 175 between the two tests with a group of 46 inspectors in an aircraft plant (presumably a homogeneous group) and one of 37 for 40 aircraft factory foremen and assistant foremen The relationship with the A C E Psychological Examination reported in the manual for 212 technical-high school seniors (apparently in Springfield, Mass) is 55, working with 230 Merchant Marine Cadets, Traxler (863) found a correlation of 37 For the L score the coefficient was 34, and that for the Q score was 26 This tendency for verbal intelligence to be at least as closely related to mechanical comprehension as quantitative intelligence is confirmed by Carnegie Mental Ability Test data reported in the manual $r = 54$ for the L score, 52 for the Q score, the subjects being 131 defense trainees It seems that in fairly heterogeneous groups abstract mental ability is moderately related to mechanical comprehension (as indeed the term implies), whereas in homogeneous groups it is quite distinct This makes its measurement in technical training institutions which select largely on an intellectual basis especially pertinent, assuming that the test actually has predictive value

Manual dexterity tests which have been correlated with Bennett's include the Psychological Corporation's Large Hand Tool Dexterity Test (disassembly and assembly of nuts, washers, and bolts with wrench and screw driver), the Minnesota Manual Dexterity Test, and the O'Connor Finger and Tweezer Dexterity Tests The first study is reported in the manual, the subjects being 89 veterans in a guidance center and 1109 paper-bag factory workers, the correlations equalled 39 and 28 The Minnesota Manual (Placing and Turning) Tests and the O'Connor tests were used by Jacobsen (396) in a study described in an earlier chapter, for 90 mechanic learners he found correlations of 21 and 14 with Placing and Turning Tests, and of -04 and 14 with Finger and Tweezer Dexterity, respectively It seems surprising that there should be a relationship between mechanical comprehension and gross manual dexterity as measured by the hand-tool test but not as measured by an arm-and-hand movement test It would seem more logical that there be no relationship at all between dexterity and comprehension, as suggested by Jacobsen's data More evidence is needed

Mechanical aptitude has been measured by other tests and correlated

with the Bennett in several studies reported in the manual and in two other studies by Sartain (669) and by McDaniel and Reynolds (493) The test reported on in the manual is the MacQuarrie, administered to 136 applicants for WPA mechanical courses and to 220 applicants for apprentice courses, with correlations of 40 and 48 Sartain's correlation coefficient was 44 for aircraft factory inspectors, McDaniel and Reynolds' was 55 for 147 defense-training students These correlations are only to be expected, in view of the use of the MacQuarrie as part of the internal consistency criterion in selecting items for the Bennett test

Spatial visualization tests used with Bennett's and throwing some light on what it measures are the Revised Minnesota Paper Form Board and the Crawford Spatial Relations Test Correlations reported for the former in the manual are consistent, ranging from 44 for 206 technical high school seniors to 59 for 136 applicants for WPA mechanical courses, Traxler (863) reported one of 39, but Sartain (669,671) found relationships of 27 and 31, and Jacobsen (396) reported a coefficient of 00 These inconsistencies are difficult to explain, but Jacobsen's finding is so unlike the others that it may perhaps be disregarded The trend then rather clearly is for the two tests to be moderately closely related, as they should be in view of the use of the Paper Form Board in selecting Bennett items Jacobsen is the only author who has reported on the relationship of the Bennett to the Crawford test, his r being only 18

Interest was correlated with Bennett scores by Moore (530) who used Strong's Vocational Interest Blank as a measure of interest His subjects were two groups of engineering defense training students, numbering 205 and 292 respectively The correlations between the Bennett and Strong's Engineering key were 30 and 35 for the two groups, for the Aviator key they were 21 and 26 for the Production Manager key they were 12 and 08, and for Carpenter they were 06 and 12 These findings suggest that the higher the level of mechanical comprehension, the higher the level of technical interest, for the higher correlations are for the technical occupations This is not confirmed by the somewhat different mechanical and scientific keys of the Kuder Preference Record (671), the correlations with which are only 15 and 15 for a more homogeneous group of foremen

Grades in technical courses, standing on examinations in technical subjects, ratings of students and learners by instructors, and ability to complete technical training courses have been used as criteria in training situations Grades made by 1834 defense industry trainees in a chemistry

course were correlated with Bennett scores by Moore, r being .36. For 137 shop trainees of Pan-American Airways the correlation with shop grades was 62 Moore also obtained correlations of 39 with final examination scores in defense-training chemistry courses, and 52 with final examinations in the physics course The latter examination was a Co-operative Test Service physics test, the manual also reports a correlation of 42 between the Bennett and College Entrance Board Physics Examination scores of 275 applicants for an engineering school

Not reported in the manual are two studies of the test's value in predicting ratings of the mechanical promise of war industry trainees McDaniel and Reynolds (493) used a group of high school students and defense industry trainees, 147 in number Their criterion was instructors' ratings of learning aptitude, speed and accuracy in acquiring muscular and manipulative skills, quality and precision of work, and eagerness in getting at the job and staying with it, combined into one overall rating of promise Ten-point scales (too refined for use by non-psychologically trained raters) with behavior descriptions were used for each of the four traits No data are presented as to the reliability of the ratings, their correlation with Bennett scores was 24, approximately that for the O'Rourke and slightly higher than those for various parts of the MacQuarrie Mechanical Aptitude Test

Jacobsen's study (396) has been described in connection with other tests He found that the correlations between Bennett scores and ratings of fitness for mechanical work as judged in courses in aircraft instruments, airplane engines, aeronautical repair mechanics, machine shop, and aircraft electricity were 35, 11, 30, 35, and 41 respectively (P E equalled 07 to 09) When combined with other tests the multiple correlations ranged from 46 (repair mechanics) to 64 (instruments), except for ratings in the course in aircraft engines, perhaps this was due to defects in the ratings in this course, rather than to differences in the psychological demands which it made on the learners

Bennett points out in his manual that many validity coefficients were obtained for his test or for very close copies of it in the armed forces One part of the Army Air Force Qualifying Examination (195), consisted of from 15 to 60, generally 30, Bennett-type items, validity coefficients for various forms correlated with success-failure in primary pilot training ranged from 14 to 38, for graduation-elimination in navigator training the validities ranged from 22 to 45, and for bombardier training, the criterion of success for which was not satisfactory, the one valid-

ity coefficient reported was 13 In an experimental group of 1080 cadets sent to pilot training regardless of test scores (214), the validity coefficient for the mechanical comprehension part of the Qualifying Examination was 47 (graduation criterion), the Mechanical Principles Test of the regular cadet test battery had a validity coefficient of 43 for this group, only two tests had higher predictive values, one entitled Instrument Comprehension II ($r = 48$) and the other a Test of General Information scored for pilots ($r = 51$)

The test was used by the Army and Navy, in various forms, for the selection of trainees in other specialties The Army Mechanical Aptitude Test included 22 items from Bennett s Form AA, plus others resembling it, the Navy had its own forms also Validity data for some of these are reported by Fredericksen (273) and by Stuit (785), but will not be cited in detail here, as the Air Force data illustrate them As Bennett's manual puts it, whenever the ability to understand machines is important the test and its derivatives are likely to have fairly high validity Navy technical courses for which the Bennett-type tests were validated are listed in Table 18, with validities

TABLE 18

RELATIONSHIP BETWEEN BENNETT SCORES AND NAVY GRADES

Course	*r*
Submarine School	
Torpedoes	23
Communications	23
Submarines	23
Engineering	39
Indoctrination School	
Seamanship	28
Ordnance	29
Navigation	36
Final Average	35

Success on the job as measured by ratings of supervisors has been correlated with Bennett scores by Bennett and Fear (70), McMurry and Johnson (500), Saltain (669,671), Schultz and Barnabas (682), and Shuman (716,717) In Bennett and Fear's study 60 machine-tool-operator trainees were tested prior to training and were rated by their supervisors for performance on the job several months later The reliability of the criterion was apparently not checked Test scores and ratings of job performance had a correlation of 64, an unusually high validity for one test which would need to be confirmed in other similar studies before

being accepted (Shuman's study, discussed below, found an r of 44 for machine operators) As a result of this finding only applicants who rated A or B on a combination of this and one other test were employed, the group as a whole making good employment records, as evidenced by the fact that, "of all new men hired since tests were installed 76 percent were rated as 'excellent' or 'good' on the job Only 8 percent were rated 'below average' None were rated as 'poor' Not a single new man, hired since tests were introduced as part of the selection procedure, has had to be dismissed because of lack of ability to do the job" Perhaps this conclusion needs to be qualified by a reminder of the facts that supervisors are generally reluctant to use the "poor" rating, and that during the war employers were reluctant to release employees

Further confirmation is provided by McMurry and Johnson, who tested 769 ordnance factory employees at the time of selection with a battery including the Bennett test Supervisors' ratings of 587 of these were obtained after they had been on the job some time Validity coefficients were computed for occupationally homogeneous subgroups For a group of 33 cranemen the Bennett test had a validity of 65, other occupational groups were also tested, but validities are not reported for the Bennett alone

Sartain's first study has been discussed elsewhere, his ratings, it will be remembered, were of performance in a refresher training course for aircraft factory inspectors already on the job whose job performance was known to the instructors The correlation between Bennett scores and this mixed training-job criterion was 32, lower than those of 65, 64, and 47 for the MacQuarrie, Otis, and Minnesota Paper Form Board In his second study, the subjects of which were 40 aircraft factory foremen and assistant foremen rated by their supervisors, the correlation between Bennett scores and ratings was — 15 This may prove that foremen in this plant were judged more by success in handling employees than by success in coping with mechanical problems, and is probably no indication of the validity of the test for mechanical and technical work Shuman's study, discussed below, suggests that in some situations the mechanical comprehension of foremen is considered by raters, Schultz and Barnabas' investigation also bears on this point

Employee relations and "budget-control efficiency" of 30 foremen and assistant foremen were rated by supervisors in the study reported by Schultz and Barnabas The foremen were tested with a battery made up of the Bennett Mechanical Comprehension Test, the Strong Vocational

Interest Blank (scored for Production Manager and Occupational Level), and the Bernreuter Personality Inventory (combined scores) The reliability of the ratings was determined by re-rating at the end of a five-month period The correlation between the combined ratings on "employee relations" and "budget-control efficiency" for the first and second ratings was 85 When T scores for the three predictors were combined a correlation of 52 with combined ratings was obtained The correlation between Bennett scores and the criterion was 11

In Shuman's study of aircraft-engine and propeller factory workers the criterion of success was supervisors' ratings of efficiency of the job Workers were rated as good, average, or poor, in consultation with rating experts In two departments the ratings thus made were correlated with ratings made by a departmental instructor trained in rating techniques The reliabilities thus obtained were 91 for 42 production engine testers and 705 for 36 inspectors, the former correlation being so high as to make one wonder about possible contamination of data through discussion by supervisor and instructor Tests were administered to operators who had been on the job for six months or more, ratings were secured after testing New applicants were also tested at the time of application, and those employed were followed up six months later and rated These two groups were combined, the possible differential effects of pre- and post-hiring testing apparently not being investigated The numbers in each occupational group varied from 25 (job setters) to 99 (foremen) Biserial coefficients of correlation were computed between the tests used (Otis, Minnesota Paper Form Board, and Bennett) and supervisors' ratings, by occupation Data for the Bennett Test are presented in Table 19

TABLE 19

BISERIAL CORRELATIONS BETWEEN JOB RATINGS AND BENNETT SCORES

Job	N	r bis	Critical Scores		Percent Improvement	
			Male	Female	Male	Female
Inspectors	49	665	34	19	12	28
Engine testers	45	17	33		10	
Machine Operators	81	44	27	18	22	12
Foremen	99	465	30		10	
Job setters	25	73	36		47	
Toolmaker learners	64	46	36		5	
Mean	363	52			18	

We have already seen, in the discussion of the Otis test, that the latter had substantial validity for all of these jobs (r = 39 to 57), it is interesting that the validities for the Bennett are lower in some cases (engine

testers) and higher in others (inspectors and job setters) It would be helpful, in such a case, to have job descriptions which would throw light on the reasons for these differences, but presumably in the engine tester's work there is no advantage in having more than the minimum required degree of mechanical comprehension (perhaps it is more a matter of manual dexterity in making connections and perceptual ability in reading dials), while in the job setter's and inspector's work higher degrees of understanding of mechanical principles make for greater worker efficiency (which is understandable if the inspectors were *engine* inspectors)

Critical scores were set for each of the tests used, those for the Bennett being shown in the next-to-the-last column of Table 19 In jobs utilizing both men and women sex differences made special minima necessary, in other jobs men only were employed The difference between these minima indicate that the machine operators' work requires least mechanical comprehension (it also requires the least mental ability), and that job setters and toolmaker learners are the most highly selected groups in mechanical comprehension this is what one would expect, and can be taken as a sign of the validity of the test Foremen, for whom supervision of personnel is more crucial than mechanical aptitude, also have a lower critical score than the more technical workers, although in this situation, unlike Sartain's, mechanical comprehension does play some part in foreman success as judged by the supervisors As the final column of Table 19 shows, the hiring of workers on the basis of the established critical minimum scores would be improved by from 5 percent in the case of toolmaker learners to 47 percent in the case of job setters, with a mean hiring improvement of 18 percent for all the jobs in question The Bennett test contributed more to the improvement of selection than either of the other tests used, except possibly in the case of inspectors and toolmaker learners

Supervisory workers in three factories were studied in another investigation in which Shuman used the same battery of tests Foremen, group leaders, and job setters were rated as to production, handling of workers, housekeeping, and overall opinion by their superiors, the total usable group numbering 208 The mean correlation between Bennett scores and ratings of several groups of foremen was 55 Minimum critical scores were established for each job, that for foremen being 30, and that for group leaders 26 When data for all supervisors were combined, the percent improvement in selection of excellent workers which would have

been effected by use of the Bennett test was 18, exceeded only by the Otis' 19 percent

Occupational differences in mechanical comprehension as measured by the Bennett test are shown by Shuman's studies (716,717) and by the industrial or occupational norms reported in the manual As Shuman's basis for establishing critical scores is not described, and as he does not present data on means and sigmas, it is not possible to integrate his published findings with the industrial norms of the manual However, we have seen that according to him, job setters and toolmaker learners require more mechanical comprehension than do other skilled and semi-skilled workers in airplane-engine and propeller factories, and that machine operators require least The critical scores (apparently close to Q1) for toolmaker learners and job setters are at the 30th percentile for trainees in an airplane factory, and at the 50th for candidates for police and fire department appointments, as shown in the manual The critical score for inspectors was at about the 23rd and 43rd percentiles when compared to the same groups That for machine operators was at the 17th and 20th These data suggest that the most skilled jobs in an airplane factory require only a modicum of such ability In the norms provided by the manual, it is the candidates for engineering positions (average education equalled two years beyond high school) who ranked first, trainees in an airplane factory second, and men in defense training courses and applying for leadman jobs third, while candidates for WPA mechanical training courses, workers in a paper-bag factory, and applicants for employment as mechanics' helpers made the lowest mean scores These data are still limited to too few occupations, in too few plants, to be more than suggestive Norms for other skilled and also for professional-technical jobs should be provided at an early date

Job satisfaction has not as yet been used as a criterion for the validation of the Bennett Mechanical Comprehension Test

Use of the Bennett Mechanical Comprehension Test in Counseling and Selection The reported relationships between the Bennett and other tests make it clear that, when the group being tested is homogeneous, there is little relationship between mechanical comprehension and intelligence, since they are both abstract functions, however, it is only natural that they should appear to have some relationship when the groups concerned represent considerable spread in mental ability This test has been seen to be closer to spatial visualization, a finding which is not surprising in view of the studies which have shown that mechanical

aptitude is in reality a combination of ability to judge spatial relations, perception, and information Similarly, we have seen that Bennett scores and technical interests as measured by Strong's Blank are moderately correlated, although the relationship was found to be negligible in a more homogeneous group of men in whom interest was measured with Kuder's inventory

The effects of age and experience on Bennett scores have not been adequately studied, although we have seen that partial data throw light on some important aspects of these problems There are no data on the development of mechanical comprehension, but this is natural enough in a composite trait It has been brought out that the easier of the male forms is too easy for brighter and more mature men, that presumed cultural influences handicap women somewhat on the men's form, but that such specific and pertinent environmental influences as training in physics do not appreciably affect men's scores apparently older boys' and men's opportunities to become familiar with the objects and principles involved are sufficiently uniform in urban American culture to make the test "universally" applicable In this respect the test is probably superior to O'Rourke's

Occupational significance of Bennett's test has been made clear in a variety of ways, even though the occupational groups included in the published norms are in too many instances really pre-occupational or at best marginal As Bennett puts it in the manual, the test is likely to be of value in jobs in which understanding machines is of fundamental importance, when dealing with people or with abstract problems other tests will have greater validity Thus engineers and toolmaker learners are characterized by a high degree of mechanical comprehension as measured by this test, good machine operators tend to have more than the general population, and foremen in some situations (presumably the more technical) are found to be superior in mechanical comprehension while those in others (presumably those in which human relations are important) do not excel in this trait but are superior in other ways

In schools and colleges the test can tentatively be used with the published educational norms, but local norms should be developed as soon as possible in view of the probable inadequacies of those in the manual The test should prove valuable in *counseling* students concerning the choice of technical curricula and occupations it may be safe to generalize from the validity data and norms to say that those aiming at semiskilled machine work might be expected to make scores above the 15th per-

centile of their high school class on Form AA, those considering skilled trades above the 25th or 45th depending upon the trade, and those aspiring to engineering and related professions above the 50th percentile for their high school grade These suggested critical scores, it should be emphasized, have not been proved appropriate for these purposes they are merely those which the normative data on hand indicate might prove valid The test can also be used in the *selection* of students for technical courses, for we have seen that the test has some validity for training in such varied courses as machine shop, mechanics, physics, chemistry, and military flying In selection programs, of course, critical scores should be established on the basis of local experience and validities

In guidance centers and employment services the three forms can be used as in schools and colleges, discussed above, and in business and industry, considered in the next paragraph The main problem in such centers will be the choice of the appropriate form in the case of male clients, it should be made on the basis of an appraisal of the education and experience of the client, with regard to levels and quality of both intellectual and mechanical content

In business and industry the value of the Bennett Test of Mechanical Comprehension should be greatest in the selection of trainees for skilled technical jobs, and for semiskilled jobs in which fairly complex equipment is used and the induction period is longer than usual Local norms and cut-off scores should be developed, as conditions and requirements vary not only from job to job but also from plant to plant The findings reported in Shuman's studies indicate the value the test can have when so used Even when experienced skilled workers are being selected the test can probably be of some value if jobs being filled require versatility of skills and ability to apply them to constantly changing situations In industrial work, as in counseling, due consideration should be given to other measurable and less tangible factors, for we have seen that intelligence, interest, and personality traits also play a part in success in skilled work, sometimes, as in some foremen's jobs, a more important part than mechanical comprehension

The MacQuarrie Test for Mechanical Ability (California Test Bureau, 1925)

The MacQuarrie Test for Mechanical Ability was developed in 1925 as a rough measure of promise for mechanical and manual occupations

As pointed out in the introductory section of this chapter, it is not a test of mechanical comprehension as such, but a battery of subtests each of which was designed to measure some factor which it was believed would be important to success in mechanical and manual occupations Subtests were designed to measure spatial visualization, manual dexterity, and perceptual speed and accuracy, on the assumption that a test made up of such items would measure mechanical aptitude. The test might well be treated in the chapter on manual dexterities, insofar as some of the subtests are concerned, or in the chapter on spatial visualization, when dealing with other subtests, it is considered here because it like mechanical comprehension tests, is an attempt at an overall measure of mechanical aptitude It has been widely used and, despite defects related to its early origin and insufficient subsequent editorial work, has held its own as a very useful test of mechanical aptitude

Applicability The MacQuarrie Test was designed for use with adolescent boys and girls, apparently as a tool for selection for trade training Subsequent work has found that the items are equally applicable to adults, and adult norms and validity data have been accumulated The original norms (504) only part of which are in the current undated manual, show that scores increase each year from age 10 to age 19 or 20, the mean raw score at age 10 being 26, age 15, 57, and ages 19 and 20, 67 and 68 respectively Mitrano (533), it is true, reported that scores decreased with age in adolescence, a surprising finding until it is noted that his sample of 13- to 16-year-olds were all in 8th grade and that the oldest pupils were therefore probably the dullest members of the class and the least well motivated On the other hand, Goodman's (298) finding that scores decreased with age in a group of 329 women radio assemblers aged 16 to 64 years is not surprising r s for subtests and age ranged from $-.21$ (Location) to $-.34$ (Tracing), r for the total score and age was $-.38$ ($P E = .03$) As one might expect, younger adult subjects tend to do better on a speed test Use of appropriate norms, discussed below, is important in view of the age differences which the original adolescent norms make quite clear

Content The MacQuarrie is a booklet made up of seven subtests, the first three of which (Tracing, Tapping, and Dotting) seem on inspection to be measures of manual dexterity or eye-hand co-ordination, the next three (Copying, Location and Blocks) spatial visualization, and the last one (Pursuit) perceptual speed and accuracy Because of these differences

in content most users of this test in validation studies have preferred to treat each part separately, a judgment which will be seen to be justified by the results

Administration and Scoring This is a group test requiring about one-half hour for administration The only special precaution required is making sure that examinees turn the page when so directed at the end of each subtest, rather than working beyond the time limit This is easily controlled by beginning at once with the directions for the next practice test, but when groups of more than 25 are tested assistance is especially important Scoring is more complex than for most paper and-pencil tests, as the scorer must, for example, examine each opening in the lines of the Tracing Test to make sure that the pencil has gone through the opening without touching the sides, a little practice soon makes it possible to make these inspections very rapidly It might be noted in passing, however, that a somewhat greater degree of mechanical aptitude on the part of the test author could have resulted in machine- or stencil-scoring for the Tracing, Dotting, Location, Blocks, and Pursuit Tests, at least when the test was slightly revised at some unspecified date after 1943

Norms. The norms provided in the manual show the scores made by an unknown number of adolescents of unspecified sex at ages 10 through 16, and for "average adults" of 17 and above These are abbreviated norms, showing only the means and critical percentiles rather than the total distributions In view of the continued increase in scores from age 16 to 19 or 20 the lumping together of all persons over 16 might be questioned, unless other data showed that the sample of older adolescents was inadequate as a result of elimination in the last years of high school This has not apparently been actually demonstrated for this test, but the fact that the mean average adult score reported in the manual's table of norms is only 62, as compared with those of 63 and 68 for 17- and 20-year-olds reported in the original norms, suggests that the latter two groups may have been somewhat highly selected rather than representative More debatable, in view of the data, is the lumping together of the two sexes in these general norms, for the norms for part scores, to be discussed below, show sex differences for some subtests Finally, the failure to specify the number of cases involved in these norms is to be deplored, although it may perhaps be deduced from the old manual that the adolescents number 1000 minus the number of 17- to 20-year-

olds, and from the new manual that the adults number 2000 or more In view of the paucity of descriptive data by which one might judge the nature and adequacy of the adolescent sample it is necessary to use these norms with extreme caution

The adult norms supplied in the current manual remedy three defects of the adolescent norms they are specific as to sex, indicate the number of cases involved (1000 of each sex), and, equally important, are arranged according to subtests The significance of the sex differences is not reported, but the trend is for women to be superior in the spatial subtests and in total scores, that the women are not superior in manual dexterity is surprising, but the significance of failure to find such a difference is not clear One detail concerning age grouping raises a question although in the table of adolescent norms 16-year-olds were not included in the average adult group and made lower scores than the latter, in the table of adult norms they are included with the adults Presumably the age differences justify one treatment or the other, but not both Finally, as in the case of the adolescent norms, the sampling is not described One thousand men and an equal number of women might be reasonably representative of adults *in general,* they might represent some *one segment* of adults, such as routine clerks, quite adequately, or they might be a hodge-podge which can be considered a sample of *no particular universe* In view of the very real difficulties which complicate the establishing of adult norms, the user of psychological tests is, in the absence of detailed descriptive data concerning normative groups, justified only in assuming that the norms are based on the last-named type of sample, i e , a meaningless hodge-podge of adults Such norms can be used only with extreme caution

More meaningful but specialized norms are provided by Bingham (94 316), based on data for 124 apprentice toolmakers from 16- to 22-years-old and employed by the Scovill Manufacturing Co early in the 1930's Bingham points out that these norms, reproduced in Table 20, correspond fairly closely to the 16-year-old norms of the original manual at the mean but include relatively fewer high and low scores they were, in fact, a more homogeneous group such as one might expect to find working on one job in a plant with a well-tried selection program

Norms for a miscellaneous group of 334 14- to 16-year-olds in a sectarian guidance center and high school in Cleveland, Ohio, have been published by Tuckman (880) As he points out, these agree rather well

TABLE 20

MACQUARRIE NORMS FOR APPRENTICE TOOLMAKERS (N = 124)
(After Bingham, by permission)

Subtests									
1 Tracing	*2 Tapping*	*3 Dotting*	*4 Copying*	*5 Location*	*6 Blocks*	*7 Pursuit*	*Total score*	*Standard scale*	*Letter Grades*
59	52	28	72	39	26	38	89	7 5	A+
55	49	26	65	37	24	34	83	7 0	A−
51	46	24	58	34	22	31	78	6 5	B+
47	43	22	51	30	19	27	72	6 0	B−
43	40	20	44	25	16	24	67	5 5	C+
39	37	18	37	21	13	20	61	5 0	C−
35	34	16	30	17	10	16	56	4 5	D+
31	31	14	23	12	7	13	50	4 0	D−
27	28	12	16	7	4	9	45	3 5	E+
23	25	10	13	3	1	6	39	3 0	E−
19	22	8	12	1	0	2	34	2 5	

with MacQuarrie's, and they supplement the latter by providing norms for subtests and for each sex In the absence of national or local norms, these should prove useful

Standardization and Initial Validation There is relatively little available on the standardization and initial validation of the MacQuarrie test As pointed out in connection with the norms, the manual is quite inadequate in the provision of detailed information concerning the test, the recent revision reading as though it had been written for untrained and unsophisticated users of tests rather than for persons who are familiar with psychometrics The original article by MacQuarrie (504) gives little on the actual development of the test, although data on the reliability and validity of the final form are provided The total score was found to have correlations with intelligence which equalled 20 and 002 as measured by unidentified intelligence tests Teachers of shop courses rated the mechanical ability of their pupils, the correlation between these and the MacQuarrie scores being as high as 48 Other such correlations were obtained but not reported, as the reliability of the ratings was not satisfactory

Pupils also did some undescribed mechanical work which was rated by judges who did not know the pupils' identity, the correlations with these criteria were .32 and 81 for two different groups, but not enough detail is supplied to make possible the judging of these quite different and in one case almost unbelievable validities This report must, of course, be viewed in the light of the methods and standards of work current at the time of publication, at that time recognition of the importance of studying the criterion was much less widespread, and it was not generally realized to what extent supporting detail is needed for the interpretation of personnel studies Despite its defects, it makes amply clear the fact that this test is one of considerable promise, worthy of the further study which it has fortunately subsequently received at the hands of others

Reliability MacQuarrie (504) reported that the reliability of the subtest scores was as follows Tracing 80, Tapping 85, Dotting 74, Copying 86, Location 72, Blocks 80, and Pursuit 76 The retest reliability of the total score was more than 90 The number of cases used in computing the total reliability was 34, 80, and 250 in three different groups, the groups on which the part-score reliabilities were based are not described. The manual makes no mention of reliability

Validity We have seen that the initial validation data published by the author leave much to be desired insofar as detail is concerned, but appear promising in a general way Fortunately, a number of studies have supplemented MacQuarrie's findings

Intercorrelations of the MacQuarrie subtests have been computed by Goodman (299) in a factor analysis study, to be described in more detail below. The coefficients range from 29 between Tapping on the one hand, and Location, Blocks, and Pursuit on the other, to 55 between Tracing and Dotting. The manual dexterity subtest intercorrelations range from 47 to 55 and the spatial relations intercorrelations from 52 to 54, while these two types of subtests are intercorrelated with each other to the extent of from 29 to 44 Correlations between dexterity and perceptual tests are of the same order, but the spatial and perceptual tests intercorrelate between 44 and 48, which suggests that the distinction may be arbitrary The factor analysis throws more light on this, by revealing indeed three factors, one called visual inspection (our perceptual ability), another spatial visualization, and the third manual movement (our manual dexterity) This last factor is important in the Tracing, Tapping, and Dotting Tests, the spatial factor in the Copying, Location, and Blocks Tests, to a lesser extent in the Pursuit Test, and

to a still lesser extent in the Tracing and Dotting Tests, and the perceptual or visual inspection factor is important in the Tracing, Dotting, and Pursuit Tests Harrell (336) found the Dotting Test saturated with a dexterity factor, the Copying, Blocks, and Pursuit Tests saturated with a spatial factor The subtests are not particularly pure tests, although the three spatial tests are relatively unweighted by other measured factors, at the same time, the classification into *Spatial* (Copying, Location, Blocks), *Manual Dexterity* (Tapping), *Manual-Visual* (Tracing and Dotting), and *Visual-Manual* (Pursuit) seems warranted for interpretive purposes

Intelligence tests have been found to have correlations with the MacQuarrie which vary from 02 to 62 Horning (380) tested 25 pupils aged 12 to 15, finding a correlation of only 02 with intelligence as measured by the Terman Group Test Murphy (556) worked with 143 9th grade boys, finding no relationship between MacQuarrie and Terman Group Test scores Holcomb and Laslett (375) used the A C E Psychological Examination with 50 engineering freshmen, found an r of 305 Morgan (540) administered the MacQuarrie and Army Alpha to boys aged 13 through 16, each age-group including from 35 to 159 members, and obtained correlations of 33, 35, 39, and 16 respectively, it should perhaps be noted that the low coefficient is that based on the smallest group Pond, as reported by Bingham (94 317), found a correlation of 38 between MacQuarrie and Otis, her subjects being 83 apprentice toolmakers Finally, both Sartain (669) and Babcock and Emerson (35) obtained correlations of 62 between MacQuarrie and intelligence tests, the former using the Otis with 46 aircraft factory inspectors and the latter a vocabulary test with 300 subjects ranging in age from 14 to 28 The last-named study found that, contrary to expectation, the correlation between intelligence and MacQuarrie scores increased with age

At first glance, it seems almost hopeless to attempt to rationalize such divergent findings But if these studies are grouped according to the homogeneity of the subjects the differences in the findings seem more reconcilable The two studies reporting no relationship, it should be noted, are probably those in which the subjects were most homogeneous· pupils in a shop course and 9th grade boys Those reporting moderately high correlations also tend to be those which were fairly homogeneous engineering freshmen, high school boys by age groups, and apprentices in one company One of the investigators who reported high correlations worked with an extremely heterogeneous group of cases Babcock and

Emerson's subjects not only ranged in age from 14 to 28, but, more important still, were reached as clients of a counseling service and students in public schools In the other study, Sartain's, the heterogeneity of the adult workers studied is shown by a mean Otis score of 28 61 and a standard deviation of 9 48, equivalent (20-minute time limit) to a mean Otis I Q of 95, minus one sigma being I Q 82 and plus one sigma being 108, this suggests that, although the adult group was small and occupationally homogeneous, it was heterogeneous in aptitudes. Since it has frequently been demonstrated that the greater the heterogeneity of the group the greater the correlation between their scores on any two psychological tests, it may probably be concluded that the MacQuarrie Test of Mechanical Ability is relatively independent of intelligence in persons of similar status, but somewhat associated with it in groups of varied individuals

Mechanical comprehension tests which have been correlated with the MacQuarrie include the O'Rourke and the Bennett Scudder and Raubenheimer (686) found no relationship (01) between O'Rourke and MacQuarrie scores, using data from 114 7th and 8th grade boys Sartain's study (669) showed a correlation of 20 between the two tests, his subjects being 46 inspectors McDaniel and Reynolds (493) reported a correlation of 51 based on 147 students in high school and defense-training courses The differences in results again appear to be due to degrees of heterogeneity in the groups, the first being probably the most homogeneous and the last undoubtedly the most heterogeneous Similar data are available for the Stenquist Mechanical Assembly Test, Scudder and Raubenheimer (686) reporting a correlation of 01 and Harrell (334) one of 61 For Bennett s test the results are more consistent Bennett (68) reports correlations of 40 and 48 based on 130 WPA and 220 apprentice training applicants McDaniel and Reynolds also found a correlation of .48 with 147 high school and defense-training students, while Sartain's (669) factory inspectors yielded a correlation coefficient of 44 for the same two tests Underlying these more consistent findings is the fact, discussed elsewhere, that the MacQuarrie was a part of the criterion used to determine the selection of items for the Bennett test

Spatial visualization tests correlated with the MacQuarrie include the Revised Minnesota Paper Form Board Morgan (540) and Sartain (669) agreed in reporting correlations of about 30 to 40, although the use of total scores somewhat obscures the relationship shown in Harrell's (336) factor analysis, previously discussed The correlation between Mac-

Quarrie Copying and the Minnesota Paper Form Board, for example, is 49 (556), showing the greater importance of spatial visualization in the Copying than in some of the other subtests

Interest in technical subjects as measured by Strong's Engineering key was related to MacQuarrie scores in one study, in which Holcomb and Laslett (375) tested engineering freshmen The relationship was lower than in the case of experience-affected mechanical comprehension tests, being only .22

Grades have been used as a criterion of the validity of the MacQuarrie in junior high schools, technical schools, engineering colleges, dental and nursing schools, and commercial schools and colleges Horning (380) had 25 boys aged 12 to 15 graded on the basis of a project completed in a shop course, and on the basis of time taken to complete the project, the correlations with test scores were respectively 79 and 72, both remarkably high Scudder and Raubenheimer (686) made a study using the grades of 114 7th and 8th grade boys as a criterion which did not, however, agree with this their validity coefficient was only 08 Unfortunately, both studies are so sketchily reported as to make evaluation difficult

Class standing achieved in technical and *industrial schools* by boys 13 to 16 was the criterion employed by Morgan (540), with from 35 to 159 boys in each age group His multiple R was 60, that for the MacQuarrie alone was not given The 147 high school and defense-training students studied by McDaniel and Reynolds (493) were rated for mechanical aptitude by their instructors and subtest validity coefficients were calculated These were as shown in Table 21

TABLE 21

CORRELATION BETWEEN MACQUARRIE SUBTESTS AND INSTRUCTORS' RATINGS

MacQuarrie	*Ratings*
Tracing	22
Tapping	− 17
Dotting	22
Copying	21
Location	10
Blocks	22
Pursuit	12
Total	14

These are certainly not impressive, that this may be due to defects in the criterion rather than in the test is a truism which the authors seem to have forgotten, for there is no discussion in the paper of the reliability of their criterion, and such ratings are notoriously unreliable That this

may be the explanation in this case is suggested by the equally low validities of the other tests used, although the multiple correlation coefficient based upon the MacQuarrie and O'Rourke subtests and the Bennett total score was .45

Grades in courses taken by aviation mechanic trainees in the Army Air Forces prior to World War II were correlated with MacQuarrie and other test scores by Harrell and Faubion (339) The correlation between this one test and grades in drafting and blueprint reading was 47

Engineering grades during the freshman year and over the whole four years of college were correlated with the MacQuarrie by Brush (122) in a study of more than 100 men at the University of Maine, the correlations were respectively 25 and 22, with probable errors of 06 The best subtest correlations were, as might be expected, those which measure spatial visualization, but these also were low, ranging from 24 to 265 for freshmen grades and from 18 to 27 for four-year marks Revised Minnesota Paper Form Board scores, on the other hand, had validities of 42 and 43 Brush cites an unpublished study by Horton in which the MacQuarrie yielded a correlation of 44 with engineering drawing grades, subtest scores ranging from 13 to 40 Equally good results were obtained by Holcomb and Laslett (375), who found a correlation of 48 between MacQuarrie scores and grades of 50 freshman engineers The discrepancies are difficult to explain, however, Brush's numbers were greater and his criterion went beyond first-year grades

Grades in *dental* schools were correlated with MacQuarrie scores in studies by Thompson (824) and by Robinson and Bellows (634) In the latter the correlations were 35 and 48 for two different groups of freshmen, 40 for sophomores In the former, the correlation with freshman theory grades was 05 (N = 158) and with practicum grades it was 11 For seniors (N = 66) the coefficients were 17 and 13 Correlations between part scores and criterion were no better for theory courses, but that between manual dexterity subtest scores of seniors and practicum grades was 32 and that between spatial subtest scores and senior practicum grades was − 27 It is noteworthy that the same trend held for freshman practicum grades (22 and − 23), and that the correlations were reliable even though slightly lower Just why the spatial parts of the test should be negatively correlated with laboratory grades is difficult to understand, although Thompson considers it logical, and the failure to confirm Robinson and Bellows' results for grades in general is also a topic for further investigation It is perhaps relevant that Sartain (669) obtained results

rather like Thompson's for manual dexterity and spatial parts of the MacQuarrie and average grades during the first six months of *nursing* training, with the difference that both coefficients were positive (.26 and .36), as logical analysis of the test and of the tasks involved would lead one to expect, for spatial judgments are important in both theoretical and practical aspects of the sciences

The predictive value of the MacQuarrie for *clerical* training in which manual dexterity might be considered important has been ascertained in several studies Using 124 entering commercial high school girls as his subjects, 37 of whom graduated three years later, Klugman (434) found the latter superior on MacQuarrie Blocks and Tapping subtests, with somewhat better scores on other subtests the differences for which were not clearly significant Gottsdanker (303) tested 51 women students in a business college, and used examinations in work with machine calculators as his criterion of success The three dexterity tests had the following validities Tapping .25, Dotting .21, and Tracing .08 The validities are such as might be expected from the nature of the tests Barrett (46) also worked with college age women, but hers were liberal arts students, 96 of whom were studying typing and 75 shorthand Final grades were the criterion No correlation coefficients were computed, but instead the effectiveness of the tests in differentiating superior from inferior students was ascertained For typing the best subtests and their critical scores were the Tracing 50, Dotting 22, and Pursuit Tests 22, for shorthand, the Pursuit Test 24 It seems odd that the Tapping Test was not also valid for typing, but it did not differentiate between good and poor typing students, the Tapping, Dotting, Copying, and Blocks Tests also had some discriminating value for shorthand, but not sufficient to justify using them in addition to the other tests which had proved more useful On logical grounds, the Pursuit Test should have the most validity, for it seems to involve to a high degree the smooth-flowing and precise coordination of hand and eye which is required in writing shorthand

Success on the job, it is interesting to note, was not used as a criterion of the validity of the MacQuarrie Test for Mechanical Aptitude until more than ten years after its publication Harrell administered it to loom fixers (334), then the United States Employment Service used it in its studies of occupational ability patterns (750), subsequent studies have been published by Blum (104), Sartain (669), and Goodman (298,300).

In his study of loom fixers Harrell used 45 subjects employed in one Southern plant, with ratings by supervisors as the criterion of success

Each employee was rated by three or four persons on a six-point scale for mechanical ability The reliability of the ratings was not ascertained, and no validity coefficient was published for this test

Sartain, as has been seen in another context, worked with 46 aircraft-factory inspectors in a refresher training course, using ratings as a criterion The correlation between MacQuarrie scores and ratings was 65, and was this high partly, no doubt, because of the greater importance of abstract abilities such as spatial visualization in training courses than in actual work Ghiselli (286) studied another group of 26 inspectors, but these were girls who inspected and packed pharmaceutical products the study has already been described (p 179) In this case the criterion was ratings of performance on the job, and the correlation with the MacQuarrie test was only 19, the lowest r obtained

Sewing-machine operators were tested by Blum (104), who selected the 25 highest-earning and 25 lowest-earning workers on piece work, using a combination of ratings and earnings as a criterion The Tracing Test was the best single subtest (not confirmed by Stead and Shartle, as discussed below), better than any other and better than the total score A critical score of 30 was established for this subtest, and would have eliminated 76 percent of the poor and 40 percent of the good operators when applied to this same group of workers Failure to cross-validate was a defect in this study, as there would certainly be some shrinkage in discriminating power Although the percentage cited would, if it remained the same in future samples, improve selection appreciably, the critical score eliminated so many successful workers that it could be applied only in an employer's market (It would have eliminated about 55 and 70 percent of two USES samples, discussed below) Other tests should be added in such a program in order to cut down the percentages of false-positives and false-negatives

In a recent thorough study Goodman (298) administered the MacQuarrie to 329 women radio assembly operators immediately after they were hired Their age range was 16 to 64, with a mean of 27, 34 percent being under 19 years of age and only 15 percent over 50 years old The job was described as follows in the job summary "Assembles radio components such as tube sockets, transformers and capacitators on chassis to form a complete set, assembles terminal boards and other small assemblies using hand tools, mounts subassemblies on chassis and secures them in place using nuts and bolts or soldering iron and rosincore solder, removes insulation from wires using sandpaper and emery cloth, and

tins stripped leads, may specialize in one phase of assembly details"

The criterion was a rating of each new employee by the vestibule-training-school instructor after the construction of three models, ratings were based on the amount of work done during a fixed period of time, and on qualitative factors such as excess or deficiency of solder and looseness of joints No check was made on the reliability of the ratings, perhaps because of operating problems, but the distribution of ratings was found to be normal after proper statistical treatment Validity coefficients for the part scores of the MacQuarrie are shown in Table 22

TABLE 22

CORRELATIONS BETWLEN THE MACQUARRIE TEST AND RATINGS OF ASSEMBLY WORK
(N = 329)

MacQuarrie Subtest	*r Ratings*
Tracing	32
Tapping	18
Dotting	13
Copying	31
Location	35
Blocks	32
Pursuit	27
Total	42

It will be noted that the validity of the total score is greater for this job than is that of any subtest, although this is not true of certain other jobs or training courses The reason for this is made clear by the fact that five of the subtests have moderate validities apparently the work is of a type which requires manual, spatial, and perceptual aptitudes rather than just one of these abilities It is because of its tapping of these three widely applicable aptitudes that the MacQuarrie has so often proved to have some validity, although other and better measures of any one of these aptitudes usually prove more valid when relevant It is worth noting that when the most effective combination of the subtests was made, the multiple R (all subtests) was 46, only four points higher than the zero-order correlation of the total score

Unlike most publishers of such studies, Goodman went further in order to ascertain the efficiency of this test in employee selection His R of 46, evaluated by means of the coefficient of alienation, shows that use of the MacQuarrie would improve the selection of radio assembly operators in that plant by about 12 percent over and above what it would be without the test The company then planned to apply the Taylor-Russell selection-ratio tables (812), selecting for employment only the top

30 percent of the distribution on the MacQuarrie It was estimated that the old method resulted in the selection of employees, 50 percent of whom were satisfactory With an R of .46, the selection ratio set at 30 percent, the Taylor-Russell Tables indicated that 71 percent of those selected with the aid of the MacQuarrie should be satisfactory At this point the wartime shortage of personnel became so acute that every applicant seeking work had to be hired, it was still possible to make such a study in retrospect, however, using procedures which it had been planned to apply to future employees The results were reported in a third article (300) Of the original 329 employees, 193 or 58 percent had left the company, 35 of these were discharged, largely for "inability to do the work" An attempt to establish critical scores was considered a failure, but those who left of their own accord made significantly better scores than those who were discharged ($M = 48$, 40), and those who remained made intermediate scores which tended to be better than those of the dischargees ($M = 44$, $CR = 1.41$) If the Taylor-Russell ratio had been used, significantly fewer dischargees would have been selected, but almost proportionately fewer long-tenure workers also would have been accepted The test did not, therefore, contribute materially to selection

The Division of Occupational Analysis of the United States Employment Service used the MacQuarrie test in its test development work, including it in the research batteries for a variety of occupations according to hypotheses suggested by analysis of the test and of the job (Dvorak in 750 Ch 6) The result was the finding that some of the subtests are valid for clerical occupations as well as for some mechanical jobs, just as one might expect in the case of tests of manual dexterity and of perceptual ability A group of 227 clerical workers were compared with 78 manual workers (not otherwise described), and were found to equal or exceed Q3 of the latter group on the Tapping, Dotting, Copying, Location, and Blocks subtests The last three may have been due to differences in mental ability, since spatial visualization is an abstract function, but the first two have been seen to be primarily dexterity tests Validity coefficients for the occupations concerned are presented in Table 23, data on occupational differences are discussed subsequently

Outstanding in this table are three facts the validity of some of the subtests for occupations in both clerical and manual fields, the unreliability of even some high correlation coefficients when checked on another sample of workers in the same job, and the different validities of tests saturated with identical factors Illustrative of the former point is

TABLE 23

VALIDITY COEFFICIENTS AND CRITERIA FOR THE MACQUARRIE SUBTESTS

(After Stead and Shartle)

Occupation	*N*	*Criterion*	*r Subtest*						
Clerical Occupations			I	II	III	IV	V	VI	VII
Card-Punch-Machine Op	121	Output	.16	12	27	05	03	− 03	13
Card-Punch-Machine Op	113	Output	05	25	19	24	07	− 04	10
Index Clerk	52	Error ratio	− 09	− 29	− 05	− 08	− 14	07	− 25
Toll-Bill Clerk	19	Output	− 10	− 27	− 24	− 04	02	07	− 28
Calculator Operator	80	Worksample	*42*	13	07	*38*	*33*	*36*	*43*
Adding-Machine Operator	26	Worksample	*38*	29	18	06	*31*	06	12
Manual Occupations									
Pull-Socket Assembler	16	Output	*38*	− 01	22	*30*	02	− 14	14
Put-in-Coil Girl	18	% effic *	20	− 29	− 28	− 24	− *40*	− 06	− 09
Power-Sewing-Machine Operator	46	% effic	27	09	29	00	17	*32*	10
Power-Sewing-Machine Operator	23	Unknown	17	12	05	20	15	*36*	*51*
Lamp-Shade Sewer	19	Output	16	− 25	− 19	− 08	*33*	*37*	01
Merchandise Packer	30	% effic	− 18	26	*31*	− 01	21	15	12
Can Packer	43	Output	18	24	*33*	20	09	− 11	09

* Ratio of time set by time study to complete work to actual time required by worker in question to complete work

the Tracing Test, moderately valid for calculating and adding-machine operators and also for pull-socket assemblers, and the Location Test, which has positive validity for the two business-machine operator groups and for lamp-shade sewers but negative validity for put-in-coil girls Illustrative of the fluctuation of validity coefficients when the samples are small are the correlations of 51 and 10 for two groups of power-sewing-machine operators, difference which might, however, be due to differences in the criteria, one of which is not specified The third fact is illustrated by the validity of the first dexterity test (Tracing) for three occupations and the doubtful validity of the second test of manual dexterity for any of the fields in question, and also by the validity of the first spatial test (Copying) but not of the second (Location) for pull-socket assemblers

Despite these discrepancies, inspection of the table suggests that there is a tendency for the Tapping and Dotting Tests, and for the Copying and Location Tests, to agree The dexterity tests tend to have some validity for various types of office-machine operators and for packers, both of which agree reasonably well with logical analysis of the tasks, the latter, or spatial tests, tend to have some validity for office-machine oper-

ators and for machine and hand-sewers It is unfortunate, since the test was designed as a test of mechanical aptitude, that no mechanical occupations were included, the differential validities for other types of occupations are helpful as indicators of possibly worthwhile groups on which to try the test for selection purposes, they are not, however, clear-cut enough to provide very helpful data in counseling This fact will be brought out especially by the data on occupational differences, for some of the high-scoring occupations are those for which low validities were reported, and some of those for which the subtests have moderately high validities are fields, the mean scores of which are relatively low—an apparent paradox which will be discussed in subsequent paragraphs

Occupational differences have apparently not been studied as such by means of the MacQuarrie Test for Mechanical Ability, but data on differences between a few jobs have been reported in Stead and Shartle (750 - 232–235) in the form of graphs which show the approximate means and standard deviations The groups of workers making high scores on the manual dexterity subtests include index clerks, put-in-coil girls, card-punch-machine operators, and toll-bill clerks, power-sewing-machine operators, can packers, and adding and calculating-machine operators tend to make low scores on one or more dexterity tests On the spatial tests those tending to make high scores were card-punch-machine operators, index clerks, and toll-bill clerks, although the can packers included many high scorers on the one three-dimensional subtest (Blocks), as did also the merchandise packers Low scores on the spatial tests were most frequently made by power-sewing-machine operators The Pursuit Test, which is both perceptual and spatial, is one on which card-punch-machine operators and electrical-assembly workers tend to make high scores, the power-sewing-machine and adding and calculating-machine operators being low

It is interesting to note that the data on occupational differences do not always agree with those on the correlation between scores on these tests and output For example, the correlation between Location Test scores and card-punch-machine operation has been seen to be 03 and 07 for two samples, while in contrast with this negligible relationship we have also seen that card-punch-machine operators make higher scores on the Location Test than most of the other groups of workers tested At first this seems inconsistent, but on second thought it is not illogical for a job to require a fairly high degree of a given aptitude, natural selection discouraging or eliminating those who lack it, and yet not to be so

dependent on it that those who possess it in a high degree excel in the work We have already seen this in connection with intelligence tests, the data showing that in many occupations the workers must have more than a critical minimum of mental ability and that additional increments do not affect success, other factors then becoming much more important So, apparently, it is in the case of other aptitude tests This means that evaluations of the effectiveness of tests in personnel selection and guidance should not be based on correlation coefficients alone

It is also true that in some low-scoring groups the correlation with success is moderately high Can packers, for example, made a relatively low mean score on the Dotting Test, but the correlation with output in their case was found to be moderately high (33) To make this point in another way, a high correlation between test scores and success does not necessarily mean a high critical minimum for employment, and a high critical minimum for employment does not necessarily mean that a high correlation will be found between the scores of unselected *workers* and success, although it would mean a substantial correlation between test scores and success in an unselected group of *applicants* for work

Job satisfaction, in the case of the MacQuarrie as in that of most other tests, has not been used as a criterion of success

Use of the MacQuarrie Test of Mechanical Ability in Counseling and Selection The evidence reviewed in the preceding pages makes it clear that the MacQuarrie Test of Mechanical Ability measures three different aptitudes manual dexterity, spatial visualization, and perceptual speed and accuracy Although some of the subtests appear to be relatively pure measures of one single factor (the Copying, Location, and Block Tests measure spatial visualization and Tapping measures dexterity), others are measures of combinations of factors (Tracing and Dotting are manual-perceptual, and Pursuit is perceptual-manual) This being the case, it was not surprising that the educational and occupational significance of the test was sometimes obscured by the use of total scores, and the significance of the subtests was found to vary with the occupation

The effects of maturation on the MacQuarrie Test appear to be an increase in scores during adolescence, followed by the decrease with later adulthood which is usually found in scores on tests in which speed is a factor Although these tendencies have been made sufficiently clear to be considered in counseling, they have not been studied in great enough detail to make possible the establishment of special norms for use in

counseling either in early adolescents or older adults in terms of status in comparison with adult workers In such cases it is possible only to use age norms (adolescents) or general adult norms (making allowances for age on a rule-of-thumb basis)

Occupations for which the test has validity include business-machine operators (calculating machines, adding machines, card-punch machines, etc), small-assembly workers (radios, electrical pull-sockets, etc), and packers (merchandise and cans), although some subtests are valid for some and not for others of these Superior aircraft factory inspectors tend to make higher total scores than less successful inspectors, and efficient radio-assembly operators surpass less efficient operators on the total score, because both jobs seem to require the combination of aptitudes represented by high total scores On the other hand, good can packers excel on some of the manual dexterity and on the three dimensional tests, but not on others, and good power-sewing-machine operators make higher scores on the Blocks Test than do inferior operators while the Copying Test has little validity for this group

School and college use of the MacQuarrie can be varied The test is useful in counseling students concerning the choice of trade, technical, and dental curricula, although its validity is not as great as some studies suggest and part-scores should be used in fields such as dentistry, with recognition of the fact that other factors are of considerably more importance than those assessed by the MacQuarrie The dexterity and pursuit subtest scores also have bearing on success in training in typing and shorthand Because of the specificity of its part-scores, the MacQuarrie is likely to be more valuable in selection for training than in counseling concerning fields of endeavor

In *guidance centers and employment services* this test can be useful in counseling clients concerning training in the fields just listed, and in screening employment applicants who are most likely to prove successful in office-machine operation and assembly jobs

In business and industry the MacQuarrie can be a useful screen for the selection of the business-machine operators and assembly workers who have the manual dexterities and spatial aptitude which make for success in such work Because of the specific factors measured by the test and the great variations in the psychological requirements of machine-operation and assembly jobs it is important that local validities and cut-off scores be established for each subtest, rather than depending on data from other

studies As Stead and Shartle's data have shown, subtest validities sometimes vary even from one sample to another, as when the Pursuit Test yielded a validity of 51 for one sample of power-sewing-machine operators and 01 for another

The Purdue Mechanical Adaptability Test (Div of Applied Psychology, Purdue University, 1946)

The Purdue Mechanical Adaptability Test was published in 1946 as a result of work designed to produce a brief test which could be used by industrial personnel workers to measure "knack" for mechanical, electrical, and related activities It was assumed that the most effective way to do this was to measure the amount of information acquired concerning mechanical, electrical, carpentry, plumbing, and related tools, materials, and processes The test is, therefore, very similar in approach and content to the O'Rourke Mechanical Aptitude Test, previously discussed in some detail It differs in that it uses only verbal rather than both graphic and verbal items, and in that it is much briefer, consisting of only 60 items Although the only published study of the test in print at the time of writing is the original one by the test's authors (151), the instrument is briefly treated because it seems likely to become a widely used and valuable instrument

Description The 60 items in the test are divided as follows woodwork and finishing, 10 items, automotive, 17, electricity and radio, 18, machine shop, 4, plumbing, 4, sheetmetal, 2, miscellaneous, 5 These items were selected from 400 which were written to tap first-hand contact rather than principles, and to utilize 8th grade vocabulary except for technical terms The 100 best items were selected on the basis of lack of relationship to an intelligence test and internal consistency and tried out on 439 high school and college students, revised on the basis of their answers and criticisms, and administered to 364 men applying for steel mill jobs and to 98 men employed in foundries and metal products manufacturing concerns Again lack of relationship to intelligence test items and internal consistency were the criteria for evaluating items, the 60-item Form A for Men being the result The weighting of the different fields of "mechanical" work was therefore based not on judgment of the appropriate representation of the types of activity in which boys and men engage, but on the proved usefulness of various types of items in consistently measuring familiarity with tools, materials, and procedures in a variety of fields in which men and boys are customarily active The result

is an empirical rather than an *a priori* weighting which takes into account the very factors which *a priori* judgment might have considered

The test takes 15 minutes to administer, and scoring is simply a matter of counting the correct responses, doubling this sum, and adding the sum of the "don't knows" Norms given in the manual are for 667 industrial applicants, not described The article by Lawshe, Semanek, and Tiffin (454) provides norms for 1015 "industrial" men, 103 non-engineering college men, 54 engineers in non-mechanical fields, and 71 mechanical engineers The latter groups are sufficiently well described for the norms to be of some value, despite the small numbers almost all were underclassmen at Purdue, the non-engineers being majors in science, pharmacy, and physical education The industrial group is not described, although it presumably includes groups mentioned in the paper, namely industrial applicants in a steel mill and industrial applicants in an optical manufacturing plant These groups are not, however, well enough defined as to intellectual or trade level for one to be able to use them as general norms they may, for example, have been applicants for skilled jobs, or applicants for unskilled employment, or, more likely, an unknown mixture of applicants for unskilled, semiskilled, and skilled jobs

The reliability of the test, determined by the odd-even method and corrected by the Spearman-Brown formula, was found to be 84 and 80 with groups of industrial and college men (454) This is not as high as is desirable and possible in aptitude and achievement tests, although it is not too low for use, lengthening the test to 80 items with a 20-minute time limit might well prove worth while.

Validity of the test has, even in the brief period since its development, been checked in a variety of ways The correlation with *intelligence* tests was demonstrated to be low by coefficients of 32 (487 industrial employment applicants) and 17 (173 college men) with the Purdue Adaptability Test When correlated with the Otis S A scores of 25 mechanics, presumably a somewhat homogeneous group like the college students, the coefficient was 08 Although its correlation with the California Capacity, Non-Language, Test was 41, that with the Language Test was only 12 (40 apprentices) Correlation with the Bennett Mechanical Comprehension and Minnesota Paper Form Board Tests were 71 and 18 for some 30 unidentified subjects, which suggests that, as one would expect, the Purdue Test measures the informational component of mechanical comprehension rather well but does not tap spatial visualization to any great extent These findings need, however, to be confirmed by other studies

with well described samples before they can be considered conclusive

The authors (454) report no relationships with *grades,* as yet, their interest having been primarily in the industrial use of the test Correlations with *occupational criteria* are mostly rank-order coefficients based on very small groups, and so can be considered only as preliminary indications of the test's possible significance If these data are followed by more comprehensive validation studies, as the sponsorship of the test suggests it will be, this is still a good deal more evidence in favor of the test than is presented in most first editions of test manuals A group of 14 experienced mechanics in an ice company were rated by their supervisors The scatterdiagram showing ratings and scores is long and narrow, suggesting a rather high correlation (81) and a rather effective cut-off score of 90 Six time-study men in a musical instrument factory were ranked by their supervisor, the rank-order correlation with the Mechanical Adaptability Test being $.75 \pm .18$ Twelve steel mill apprentices were tested at the time of hiring and ranked by their supervisor after they had been on the job, the rank-order correlation being $.39 \pm .24$ Data for several other groups are reported, but as they were used in standardizing the test they are not meaningful

Although no data on *occupational differences* are as yet available, the authors report differences between pre-occupational college groups which are rather informative The mean scores for 71 mechanical and aeronautical engineering students were 103, civil, metallurgical, and electrical engineering students 96, and science, pharmacy, and physical education majors 92 The critical ratios between these groups were 3 8 (mechanical vs non-mechanical engineers), 6 3 (mechanical engineers vs non-engineers), and 2 1 (non-mechanical engineers vs non-engineers) These significant differences suggest that this test is indeed a mechanical rather than a scientific, or even physical science, information test, and that it should be most useful in the counseling and selection of persons considering mechanical work

As more studies are made it will be helpful to have comparisons of this test's effectiveness with that of the O'Rourke, as the most nearly similar test available, and with that of the Bennett, as one which differs from this in that it attempts to measure comprehension of principles rather than familiarity with tools and processes More detailed and specific industrial norms will be helpful in counseling, although in selection local norms must always be developed And validation studies based on larger

groups with refined criteria of success are needed in order that the occupational significance of the test may be known In view of the simplicity of the vocabulary, educational validation for trade and technical courses should not be neglected As such evidence is forthcoming the Purdue Mechanical Adaptability Test will probably become a widely used and useful diagnostic and prognostic instrument

CHAPTER XI

SPATIAL VISUALIZATION

THE ability to judge the relations of objects in space, to judge shapes and sizes, to manipulate them mentally, to visualize the effects of putting them together or of turning them over or around, is generally referred to as spatial visualization It is an aptitude which has long been considered important in such clearly similar activities as machine-shop work, carpentry, and mechanical drawing, in which the worker must judge shape and size and translate two-dimensional drawings into three-dimensional objects, and which has been considered likely to be important in certain other occupations, the principal activities of which were not quite so clearly similar, such as engineering and art

Work in the measurement of spatial judgment began, however, as one aspect of the measurement of intelligence rather than as an attempt to measure a special ability of significance in certain occupations Clinical psychologists, attempting to devise non-verbal or performance tests of intelligence which would be useful in appraising the mental ability of persons with limited formal education or whose linguistic development might in some other way have been handicapped, resorted to the familiar puzzle-type test in which the subject is required to put objects together in such a way as to make a pre-determined pattern Sometimes the pieces to be assembled were parts of a picture, as in the Mare-and Foal Test used in the Pintner-Paterson Scale of Performance Tests, in such cases the cues relied upon by the examinee are partly spatial (the shape of the curved outlines of the parts) and partly experiential (e g , the head must fit at the end of the neck) In other tests experiential content was not utilized, as in the case of the Casuist Board, in which geometric figures are put together to form large wholes, also geometric In such tests, the removal of cues based upon and requiring the analysis of experience was part of an effort to make the test truly a measure of mental ability rather than one of education As subsequent work showed, it resulted in the measurement of a trait which is related to mental ability in childhood but relatively independent in adulthood

When large-scale testing operations made it desirable to develop group tests of the performance type, Army psychologists in World War I produced Army Beta, a paper-and-pencil version of a performance scale The subtests, like those of the apparatus tests, involved completing incomplete figures of people and other familiar items in which analysis of content could help the examinee, and judging the relations of geometric figures, in which it was hoped that abstract reasoning alone would play a part. As such paper-and-pencil tests of spatial judgment were made available for adult use, form boards were also developed for use with normal adults Link developed a Form Board which subsequently developed into the Minnesota Spatial Relations Test, and Kent and Shakow devised a series of Form Boards which has models for clinical use with mental patients and for industrial use with normal adults

Because of the emphasis on the measurement of the intelligence of special groups which pervaded early work with tests of spatial relations, and the subsequent application of such tests to industrial use, students of testing are often confused by what seems to be a serious inconsistency in the use of tests by psychologists They find tests of spatial judgment figuring prominently in intelligence tests such as Army Beta, the Army General Classification Test, and the American Council on Education Psychological Examination, and also masquerading as tests of a special aptitude as in the case of the Minnesota Spatial Relations Test, the Minnesota Paper Form Board, the Kent-Shakow Form Boards, and the Blocks Test of the MacQuarrie Test of Mechanical Ability The question arises, is it possible that the same type of item can measure both intelligence and a special aptitude not related to intelligence?

The theoretical explanation of what actually seemed to be the case was slow in coming, because of the divergent interests and practical concerns of both clinical and personnel psychologists But it was implicit in data familiar to most psychologists, for it had long been known that performance tests of intelligence (i e , form boards, tests heavily saturated with spatial visualization) did not correlate well with other tests of mental ability and gave poor predictions of school achievement, increasingly so with increasing age This suggested that spatial judgment might be a special aptitude which develops at approximately the same rate as other mental abilities, and therefore provides a fair measure of mental age in childhood, but that, being a special aptitude, the degree of spatial judgment possessed in middle adolescence or adulthood is not a good indicator of the amount of any other mental ability possessed by the

individual This has since been confirmed by Garrett (281) in an analysis of test data for the appropriate years, and, in another way, by Thurstone's work (839), in which it was demonstrated that what is thought of as intelligence is, in fact, a number of special aptitudes In this analysis spatial visualization emerges as one special aptitude, distinct from the verbal, numerical, perceptual, memory, and other aptitudes which are relatively independent of each other in homogeneous groups but tend to be associated in heterogeneous groups A spatial relations test is therefore effective in classifying people according to "general" ability when wide ranges of ability are in question, and so has a part in a test such as the AGCT, on the other hand, when a group of fairly similar general ability is being studied, whether they be factory workers or college students, scores on tests of spatial relations are found to be related to success in certain types of activity without being good predictors of success in others We have already seen, for example, that verbal scores on the A C E Psychological Examination give equally good predictions of success in social studies and in mathematics, whereas quantitative scores, which are partly based on spatial items, give substantially better predictions of success in mathematics than in social studies

Of the tests which have been developed for the measurement of spatial visualization, the most widely used in vocational counseling and selection have for some years been the *Minnesota Spatial Relations Test* and the Likert-Quasha Revision of the *Minnesota Paper Form Board* These will be discussed in this chapter, it will be seen that the tests are impure, for they measure certain other factors to a lesser degree In addition to these special tests of spatial judgment the user of tests should keep in mind the spatial subtests of composite tests or test batteries such as the *Blocks Test* of the MacQuarrie Test of Mechancial Ability, the *Surface Development Test* of the Chicago Tests of Primary Mental Abilities, and the *Space Relations Test* of the Psychological Corporation's Differential Aptitude Tests, all of which are discussed elsewhere in this book

Another very well-known test of spatial visualization is Johnson O'Connor's *Wiggly Blocks* (122,341,416,626), the widespread use of which would justify discussion in this chapter if it were not so unreliable as to make it useless *Mellenbruch* (523) developed a series of similar blocks at about the same time but did little with them, and *Uhlaner* (unpublished study) has recently developed a reliable series of curved blocks which may in time prove useful further research with Uhlaner's

series should be encouraged and will bear watching *Crawford* (933) also has a test which, like the three wiggly blocks tests, attempts to measure spatial visualization with three-dimensional material, but this test also is new and there is as yet too little evidence to judge it by In view of the fact that judgment based on two-dimensional materials such as those of the Minnesota, Thurstone, and Psychological Corporation tests may not be identical with judgment of space based on three-dimensional materials (there is only one study to suggest that it is), it is to be hoped that further theoretical and occupational research will be conducted with the more reliable of these tests

The Minnesota Spatial Relations Test (Marietta Apparatus Co and Educational Test Bureau, 1930)

The Minnesota Spatial Relations Test was developed by the Mechanical Abilities Research Project of the University of Minnesota because of the promise of the Link Form Board (588) The latter test had a reliability of only 72 as determined in the preliminary work of the project, by using four boards instead of one, the new test achieved satisfactory reliability It has since been used in the Minnesota Employment Stabilization Research Institute which added valuable normative material, and in several other studies to be discussed below, but the administrative expense of apparatus tests and the fact that it has a rather good paper-and-pencil equivalent have kept it from being as widely used and studied as some other tests It is discussed here because it is a purer test of spatial judgment than the paper-and-pencil tests, as will be seen later, and therefore contributes materially to our understanding of the trait and has special value in testing for the less abstract or academic types of technical training and employment

Applicability Like the other tests of the Minnesota Mechanical Abilities Project, the Spatial Relations Test was first used with junior high school boys taking trade courses, but was designed with the objective of making it usable with older adolescents and adults Use of the test with boys as young as 11 years old and with adults of all ages has confirmed the belief that the nature of the task is such as to make it applicable to a wide range of ability, spatial judgment beginning to mature early enough for the items to be meaningful even before adolescence As the aptitude is still maturing during adolescence age norms are of course needed, and here as elsewhere a problem is encountered in the vocational counseling of adolescents If one uses age norms in interpreting the test

scores of high school students, one runs the risk of encouraging a student who is superior to his class or age group in spatial judgment to enter an occupation for which he may actually be lacking in the aptitude in question, because those who enter the occupation may be so highly selected in that respect that he is actually at the bottom of the occupational distribution even though near the top in the general norms for his age Age norms are available, as are some occupational data, but developmental conversion tables are lacking which would enable a counselor of high school students to determine how able a given boy or girl will be when adult to compete with persons engaging in various occupations Judging by the age norms, spatial visualization increases until age 14 and remains constant at ages 15, 16, and 17, there is a suggestion of an increase at age 18, the mean score for which is somewhat higher than that for the three preceding years, but the difference is not great and may be due to elimination of some of the less able in the older sample the 80th and 90th percentiles are about the same for ages 15 through 18, which fits in with the explanation of the difference on the basis of sampling

Content The Minnesota Spatial Relations Test is made up of four form boards, of which A and B use the same pieces and C and D have common parts The arrangement of the parts differs, however, in the two members of each pair, so that having placed them in Board A presumably helps one in doing Board B only by orienting one to the task and materials it does not teach one where the parts go The parts themselves are cut from a rectangular board about three feet long by one wide, there are three pieces of each shape, but of varying sizes, arranged close together but not adjacent to each other in the board The shapes include crescents, squares, angles, and odd-shaped geometrical forms

Administration and Scoring The test is administered individually and requires from 15 to 45 minutes, the average adult finishing all four boards in 20 or 25 minutes Although it is not stated in any of the publications or manuals dealing with the development or administration of the test, the subject stands while taking the test Failure to include this simple but basic detail in the manuals has resulted in the test being administered with the client seated, in some guidance centers, and standing in others, while some known to the writer have let the subject decide which way to do it At one of the latter places the staff reported that subjects concluded it was more easily done standing, despite this fact, no one took the trouble to write to the test author and ask how it was administered to the subjects on whom the norms are based!

A letter from Professor Paterson, dated August 14, 1946, states "It is the rule to have subjects stand This isn't the worst part of the story Manufacturers have substituted different kinds of materials at will with the result that the norms may not apply" In order to ascertain the possible effect of these different methods of administering the test, the writer and a colleague (Charles N Morris) conducted an experiment in which the test was administered to groups with the subjects sitting while taking all four boards, standing for all four boards, sitting for the first two but standing for the last two, and standing for the first two but sitting for the last two Although comparison of the mean gains on the second two boards over performance on the first two failed to demonstrate clearly that higher scores are made if the test is taken standing up, there was a tendency for those who took the test standing to do somewhat better than those who were seated In administering the test sitting down, then, psychometrists may be penalizing their examinees and obtaining an inadequate picture of their spatial judgment

In view of the diversity of materials which, as Paterson points out, have been used in manufacturing the test, experiments should be conducted which check up on the effects of the variations on the suitability of the test norms The point has already been made concerning the two different forms of the Minnesota Manual Dexterity or Rate of Manipulation Test In connection with the spatial test, the point might be made that wooden equipment may fit less readily than metal, and that different rates of wearing render tests made of one type of material unusable more quickly than another One manufacturer paints board and inserts black, as in the original wooden materials, another provided green-topped wooden inserts for black metal boards In the Army Air Forces Aviation Psychology Program it was found that frequently used equipment soon wore so badly that the nature of the task was considerably changed The form boards used in the experiment referred to in the preceding paragraph were not only somewhat worn, which made some pieces fit more easily, but somewhat warped, which made others fit less easily than they had originally The effect of this on test scores and the suitability of the norms has not been checked

Apart from these questions of the examinee's position and the nature of the test materials, administration of the test is straightforward Scoring, in the original work of the mechanical ability project, involved obtaining the total number of seconds required to complete all four boards, the norms for boys are based on this procedure This is the

method described in the manual published by the Educational Test Bureau, publisher of the green-topped inserts and black-painted boards The Minnesota Employment Stabilization Research Institute experimented with methods of scoring the test and found, however, that its reliability was increased by treating the first board as a practice trial and scoring only Boards B, C, and D (187) The general adult and occupational norms obtained in the MESRI work were therefore published in terms of the last three boards (306) and this is the recommended method of scoring

Norms The boys' norms provided by the Mechanical Ability Project cover ages 11 through 18, in grades 7 to 12, the numbers in any given group ranging from 55 to 150 All of them were boys in Minneapolis and St Paul schools, and while they may have been a good local sample there are no data to enable one to judge the applicability of these norms in other localities Norms are also available for 57 arts and 201 engineering college students at the University of Minnesota, all freshmen These norms are based on time for all four trials, if it is desired to use them, time for Board A should be recorded and included in the total The Board A score will not be used, however, if the norms compiled at the MESRI are utilized These are based on the now familiar standard sample of 500 employed men and women, and on various occupational groups of from 20 to 489 persons each They are available in abbreviated form in the Educational Test Bureau Manual, recomputed for all four boards, but this is an inferior method In view of the paucity of data about the norm groups in the Educational Test Bureau manual, the inferiority of its scoring system, and the relative unavailability of the Minnesota bulletins in which the better type of norms are published, general adult norms are provided in Table 24 and occupational median scores, also from the MESRI, are provided in Table 25

Standardization and Initial Validation When existing tests were being surveyed for possible use in the research of the Minnesota Mechanical Ability Project, Link's Form Board seemed one of the most promising Included in the preliminary research, it proved to have less reliability than that needed for its scores to be usable in individual diagnosis It was therefore lengthened by making a total of four boards with the same type of items, and a satisfactory reliability was obtained

Like the other tests in the Mechanical Ability Project, the Spatial Relations Test was subjected to rather thorough study and validated against success in mechanical activities It was found to have a low cor-

relation with intelligence as measured by the Otis (r = 18), the group was a fairly homogeneous one of 100 7th and 8th grade boys It had a rather high correlation with the Minnesota Mechanical Assembly Test, based on the same group (r = 56) and with the Stenquist Picture Test (r = 42) When correlated with a mechanical interest inventory a relationship was again found, r being 46 Scores were not, however, related

TABLE 24

ADULT NORMS FOR THE MINNESOTA SPATIAL RELATIONS TEST

Raw score (in seconds) *Sum for Boards B, C, and D* *Men*	*Women*	*Standard Score*	*Centile Rank*	*Letter grade*
				A+
608	605	7 0	97 7	
				A−
652	648	6 5	93 3	
				B+
726	758	6 0	84 1	
				B−
814	838	5 5	69 1	
				C+
916	937	5 0	50 0	
				C−
1047	1037	4 5	30 9	
				D+
1218	1156	4 0	15 9	
				D−
1442	1354	3 5	6 7	
				E+
1583	1571	3 0	2 3	
				E−

to the father's occupation, the household chores engaged in by the boys, and similar environmental data

Validation in this early stage was done against ratings of the quality of shop work done by the boy the work was a standard task carefully rated by the instructor The group were the same 100 7th and 8th graders The correlation of 53 showed that this was one of the most valid tests in the Minnesota battery for the prediction of success in mechanical activities

Reliability Using all four trials in computing the score, the original study of the Minnesota Spatial Relations Test yielded a reliability of 84 based upon scores of 100 7th and 8th grade boys (corrected for attenuation) When the last three boards only were counted, with Board A serving as a practice trial, the reliability for 482 adult men in a selected sample of the employed population was 91 (187)

Validity Criteria used in studying the validity of the Minnesota Spatial Relations Test include the usual variety of tests of other abilities, grades in school and college courses, ratings of work samples, and differentiation between persons in various occupations Ability of the test to yield predictions of success in employment has not been studied, perhaps

TABLE 25

MEDIAN SCORES FOR VARIOUS OCCUPATIONAL GROUPS

Number	*Group*	*Median Percentile*
102	Garage mechanics	85
170	Manual training teachers	75
62	Ornamental iron workers	69
113	Men office clerks	66
20	Draftsmen	59
29	Minor bank officials	59
84	Retail salesmen	55
47	Life insurance salesmen	55
489	Occupationally unselected men	50
26	Minor executives	46
69	Janitors	30
124	Policemen	27
33	Casual laborers	2

because of the difficulty of administering an apparatus test to large numbers of employment applicants and the availability of a paper-and-pencil version of the same test (the Revised Minnesota Paper Form Board, discussed in the next section)

We have seen that the original work with the spatial relations test yielded a correlation of 18 between spatial scores and scores on the Otis Self-Administering Test of Mental Ability In an unpublished study of 100 NYA youths aged 16 to 24 the writer obtained a correlation of 25 between the same two variables Andrew (21) correlated spatial relations test scores with scores on the Pressey intelligence tests, finding r's of 43 and 36 for groups of 334 unselected men and 131 unselected women in the MESRI project and an r of 25 based on 200 women clerical workers The higher coefficients were obtained with more heterogeneous groups such as unselected adults, and the lower figures with less heteroge-

neous groups such as 7th and 8th grade boys, it seems legitimate to conclude that in homogeneous groups there are variations in ability to visualize spatial relations which are quite independent of general mental ability, and that in heterogeneous groups the relationship between the two is positive but not high enough to make one useful by itself as a predictor of the other

Manual dexterity has been studied in relation to spatial judgment by Andrew (21) and by the writer in an unpublished study The former investigator correlated scores on the Minnesota spatial test with scores on the O'Connor Finger and Tweezer Dexterity Tests, her subjects were 200 women clerical workers The correlations of 28 and 31 showed that the two types of aptitude overlap slightly, but are virtually independent In the writer s study of 100 NYA youths aged 16 to 24 the Minnesota Manual Dexterity Test (Placing) yielded a correlation of 05 with the spatial test, confirming the findings of the original unpublished study of the placing test which was developed in order to ascertain the role of manual dexterity in the Minnesota Spatial Relations Test The conclusion that differences in manual dexterity do not affect scores on the spatial test therefore seems warranted

Mechanical comprehension was seen in the last chapter to be composed of spatial judgment and mechanical information The correlations between scores on spatial visualization tests and tests of mechanical comprehension were reviewed and discussed in some detail, and are therefore not repeated here

Spatial visualization has been measured by other instruments, the scores of which have been correlated with those on the Minnesota apparatus test The original, free-response, form of the Minnesota Paper Form Board was reported by Paterson et al (588) to have a correlation of 63 with the apparatus test In the writer's unpublished study of NYA youths, the correlation between the Revised Minnesota Paper Form Board (multiple-choice form) and the apparatus test was found to be 59, Harrell (336) found it to be 65 No data have been seen concerning relationships between scores on this two-dimensional test of spatial relations and such presumably three-dimensional tests as the Wiggly Block and the Crawford Spatial Relations Test, although it would seem to be very important to ascertain the relationship between ability to judge relationships of two-dimensional objects and ability to think in terms of three-dimensional space It may be that, in working with two-dimensional objects, one actually works in three dimensions, mentally turning objects

around and over, so that there is no real difference between the two types of tests, but this has not yet been demonstrated to be the case

Factor analysis studies including the Minnesota Spatial Relations Test have been carried out by Andrew (21), Harrell (335,336), Wittenborn (935) and the Staff of the Occupational Analysis Division of the United States Employment Service (735) Andrew's study focussed on the Minnesota Clerical Test, but her factor analysis confirmed the existence of a distinct spatial factor Harrell worked with a total of 37 variables, including the Minnesota Spatial Relations and Mechanical Assembly Tests, the MacQuarrie, the Stenquist Picture Test, and Thurstone's Primary Mental Abilities Tests He located five factors, including spatial visualization, perceptual ability, and manual agility, the first-named factor was the important one in the Minnesota Spatial Relations Test, although when *accuracy* was scored rather than *time* the perceptual factor also played an important part Wittenborn's analysis of the definitive Minnesota battery isolated only a spatial factor in the Minnesota Spatial Relations Test, this factor was found to be the only one of importance in the Paper Form Board, the Assembly Test, the Mechanical Interest Analysis Blank, and, most significant of all, the shop operations quality criterion, thus further confirming the conclusion that spatial visualization is a distinct factor and the principal factor underlying aptitude for mechanical work

The USES study, of which only a summary report has been published, found that the Minnesota Spatial Relations Test is heavily saturated with a spatial factor, and that two other factors play a part in it One of these was a space-perception factor, isolated in this study and in Harrell's but not in Andrew's or Wittenborn's, presumably because of the smaller number of tests used in the last two studies The other was difficult to define, it has a wider significance than Thurstone's induction factor, and seemed to have some of the properties of Spearman's general factor, since they used a multi-factor method of analysis the authors hesitate to call it general intelligence, but consider it more likely to be that than anything else Since the subjects used were adults, aged 17 to 39, the finding of a general intelligence factor would be important, not only because it would explain why spatial tests can be used as measures both of general ability and of a special aptitude, but also because it would contradict the theory of group factors which, in America, has been accepted to the exclusion of Spearman's two-factor theory Obviously, the USES data must be reported in more detail, and confirmed by other

studies, before conclusions of such major importance can be drawn In the meantime, it can be concluded that there is a distinct spatial factor which is the most important element in the Minnesota Spatial Relations Test and in mechanical success, and that perceptual ability does also play a part in this test

Another approach to this question is available through an unpublished thesis by Tredick (869), reported by Goodman (297) In this work Tredick correlated scores on the Minnesota Spatial Relations Test with factor scores derived from Thurstone's Primary Mental Abilities Tests administered to 113 freshman college women Significant correlations were found with the perceptual, spatial, and reasoning factors (55, 49, 47), and with the other reasoning or deductive factor (33) These data tend to confirm the USES findings in so far as components in the Minnesota test are concerned

Grades and ratings of performance in mechanical tasks were used as criteria by Brush (122), Tredick (869), Stanton (749), and Steel, Balinsky, and Lang (751) Brush used 104 engineering students at the University of Maine as his subjects, correlating spatial relations scores with freshman and four year grades, the results were disappointing, the r's in both cases being 06 It should be noted here that the Revised Minnesota Paper Form Board yielded validity coefficients of 42 and 43, which suggests that the heavier loading of intelligence in the paper version of the test makes it superior for predicting success in technical activities which are as abstract as college engineering courses Tredick also found this to be the case in a different college curriculum The students studied by Tredick were 113 freshman students of Home Economics at the Pennsylvania State College, her criteria being grades in several courses and semester-point-average for the first semester Correlations between test scores and grades were 20 for Art, 22 for Chemistry, 02 for English Composition, and 23 for semester-point-average The relationships are in the expected directions, but not high-enough to make the test usable by itself, it might have some value in a battery of unrelated tests

The nearest approach to a repetition of the original validation of the Minnesota tests was made by Stanton (749), who correlated scores on Minnesota Battery A against ratings of shop work performed by deaf boys and girls She worked with 121 boys and 36 girls, aged 12 to 14 The battery as a whole had correlations of 48 and 46 with the ratings, the validity of the spatial test alone was not given While not as high as the coefficients reported by Paterson (588) these are high enough to make the

test useful in counseling and selection when combined with other data The work sample approach was used also by Steel and associates, in a study already discussed elsewhere in this book For boys the correlation was .25, for girls, .39, as pointed out in a previous discussion, experience may have counteracted the effect of individual differences in aptitude in the boys more than in the girls, but in both cases the test had some validity

Success on the job, it has already been pointed out, has not been used as a criterion of the validity of the Minnesota Spatial Relations Test Ross (651) established a critical score for machine-tool trainees, setting it at the 30th percentile There were approximately 40 trainees But the criterion was grades in on-the-job training

Occupational differentiation on the basis of spatial relations test scores was studied first by the Minnesota Employment Stabilization Research Institute (306) and then by Teegarden (816) In the former study garage mechanics were found to make a median score equal to the 85th percentile of the general population, while manual training teachers stood at the 75th, ornamental ironworkers and men office clerks were also one sigma or more above the median (69th and 66th percentiles) Draftsmen were, surprisingly, at only the 59th percentile, the middle range included also such groups as retail salesmen, bank clerks, minor executives, and life insurance salesmen, while the lower ranges included janitors (30th percentile), policemen, and casual laborers (27th and 2nd percentiles) These differences are about as might be expected, except for the fairly high standing of the office clerks and the lower standing of the draftsmen, perhaps the latter would show up better on a paper-and-pencil test such as the Minnesota Paper Form Board, which would seem to approximate the medium in which they work more closely than does an apparatus test

The group studied by Teegarden was younger and less experienced, and her general adult norms were locally established, which makes impossible the merging of her occupational norms with those of the MESRI project without going back to the raw scores Within the limitations of her sample, it is instructive to note that there were no groups which make significantly high scores, with the exception of male operatives performing hand work in factories, who stand at the 74th percentile, and female assembly workers at the 72nd percentile But women hand operatives stand at the 55th, leading one to question the data for men, the numbers were not large, ranging from as few as 22 to 123 workers per group

Women packers and wrappers were at the 67th percentile, men at the 62nd All other groups of men and women were between the 44th and 65th percentiles As none of the occupations studied were skilled or technical occupations, the failure to find clear-cut differentiation is not surprising The MESRI occupational norms are much more helpful, we have seen that they revealed a tendency for technical and skilled workers to make high scores, and for others to make average or low scores, depending upon the intelligence level

Job satisfaction may be related to having a modicum of the ability required to perform the tasks which constitute the job, but the role of spatial visualization in vocational satisfaction has not been investigated

Use of the Minnesota Spatial Relations Test in Counseling and Selection Data reviewed and discussed in the preceding paragraphs make it clear that the Minnesota Spatial Relations Test measures at least three factors, the most important of which is ability to visualize and judge spatial relations Ability to perceive spatial differences is also tapped by the test, and indeed it is difficult to imagine a test of ability to judge spatial relations which would be entirely independent of ability to perceive spatial differences and similarities The third factor is reasoning ability, something approaching general intelligence, which plays a part in this test but is less important than the first two factors Because of the common rate of maturation and because of the fact that abstract reasoning plays a part in the test used to measure spatial judgment some relationship is found between the spatial relations and intelligence test scores of heterogeneous groups, despite this, the spatial relations test can be thought of as measuring something distinct from intelligence when working with homogeneous groups

In working with college students this means that one can expect a large percentage of average and moderately high scores, while in less able groups one will encounter more low average and low scores, these must be seen in perspective, the counselor realizing that a moderately high spatial score in a very able person does not mean special aptitude for professional-technical work and that a high average spatial score in a person of low average intelligence may well indicate promise for the skilled trades

Changes with age were seen to take place up to about age 14, after which it appears that the aptitude is relatively stable More work needs to be done before this can be considered conclusively demonstrated, but it seems a safe working principle

Occupationally viewed, the Minnesota Spatial Relations Test measures an aptitude which is found in a higher degree in workers in skilled trades and professions such as automobile repair work, manual training, and ornamental iron This is true also of workers in semi-skilled occupations in which job analysis suggests that spatial judgment should be important, these have been found to include hand-working operatives in factories, assembly workers, and packers and wrappers Although one would expect draftsmen to excel on a test such as this, the one study which included such workers found that they were only high average in spatial ability as measured by this test This seems somewhat anomalous and indicates a need for caution in making assumptions concerning the test, further studies should be made of the relationship between drafting success and scores on this test Most office and minor executive groups make moderate scores on the test, presumably because they tend to be of moderate intelligence Semi-skilled and unskilled workers in occupations not requiring spatial visualization tend to score average or below on this test, because of selection and because they tend to have less general intelligence than other workers

In schools and colleges the Spatial Relations Test is useful for selecting students who are likely to do well in shop courses, although it is of less value for the more abstract types of technical training than for the more concrete

In Guidance Centers and Employment Service Offices the test can be helpful in cases of clients considering the choice of technical occupations, especially at the semiskilled and skilled levels for which a paper form board is sometimes too abstract It has value in helping in the choice of trade and technical training, and in determining a client's prospects of making a quick adaptation to the demands of certain semiskilled jobs for which training is offered during the induction period, these latter include especially work such as assembly of vari-formed parts, machine operation, and packing objects of different shapes and sizes

Business and industrial personnel workers should find the test useful in selection of the type just described above As an aptitude test it is most useful, obviously, in selecting people for training in skilled occupations, this will happen most often in schools, but also to some extent in industry in connection with apprenticeships It can have much greater value in industry in the selection of semiskilled employees who can quickly adapt to new jobs, who can readily master procedures of machine operation or assembly, and who, because of the speed and accuracy with

which they judge size and shape, will produce more per hour of work and do it with less waste of materials

The Minnesota Paper Form Board, Likert-Quasha Revision (Psychological Corporation, 1934)

The first form of the Minnesota Paper Form Board, used in the Minnesota Mechanical Ability Project (588), was a completion test based on the Geometrical Construction subtest of Army Beta, the non-verbal intelligence test developed by the U S Army during World War I Since the scoring of completion items is laborious and subjective, requiring that the scorer scrutinize each response and make judgments as to its adequacy, it seemed highly desirable to find some way of converting the Minnesota Paper Form Board into a multiple-choice test This was done by Likert and Quasha, unfortunately not early enough to be included in the MESRI studies (617) However, the early Minnesota studies of the completion test are probably indicative of the nature of the validity of the revised test, and a variety of minor validation studies have been made with the revision

Applicability Army Beta was designed for and standardized upon unselected adults, the completion form of the Paper Form Board was developed for a study using early adolescent subjects, and the multiple-choice revision was designed for use with and standardized upon adolescents and adults The directions are simple enough for children in the upper grades, the range of difficulty of the items is such that 10-year-old boys make a median score of 22 compared with the adult male median of 34, the 5th percentile in each case being 6 and 16, indicating that individual differences are revealed at both age levels The items seem to have a reasonable amount of challenge at all age levels, despite their abstract form

The effects of maturation can be studied in two of the sets of data provided by the 1941 test manual One of these consists of the age norms for 9, 12 and 15-year-olds in the schools of Kearney, New Jersey, the other of data for grades four and five in the Bronx In the former instance a 25-minute time limit was used, instead of the usual 20-minute limit The median scores for the three age levels (boys) were 18, 32, and 38, revealing a more rapid increase in the six years from 9 to 15 (three points per annum) than in the three years from 12 to 15 (two points per annum) This suggests that the growth of this ability begins to level off early in the teens, although it does not indicate the age at which the plateau

begins The grade data confirm the changes in pre-adolescence, but go no further Lefever and others (459) found no relationship ($r = -.14$) with age in adults As in the case of the MacQuarrie, Mitrano (533) has drawn some conclusions concerning age changes which are based on spurious factors in his data, and are therefore unwarranted Studies should be made which would throw more light on the question of the age of levelling-off, to make possible the construction of developmental conversion tables such as are needed when the scores of growing individuals are to be compared to those of mature persons established in an occupation The grade norms in the 1948 manual throw no more light on age differences

Content The test consists of 64 items Each item is made up of a "stem" and five possible choices from which to select an answer The stems are the disarranged parts, from 2 to 5 in number, of a geometric figure The responses are assembled geometric figures, only one of which could be made by putting the parts of the stem figure together The problem is in each case to select the figure which corresponds to the assembled parts, which must sometimes merely be mentally pushed together in order to make an appropriate whole and sometimes mentally turned around or even over The items therefore resemble those of the real form board, except that there can be no trial and error work with the Paper Form Board all the matchings of shapes and sizes must be done mentally

Administration and Scoring The test is preceded by practice problems, with 20 minutes of working time allowed for the test proper It is necessary to demonstrate how the booklet opens, and to be sure that the many examinees who prefer to follow their own visual cues rather than the psychometrist's spoken directions do actually observe the demonstration If this is not done correctly, some booklets will be turned in with a page of easier problems skipped and some more difficult problems attempted, making scoring impossible Scoring is done by matching marked spaces with a key, is objective and simple Forms adapted to machine scoring have been published, with special norms

Norms Because of piecemeal standardization of the Likert-Quasha revision the norms for the test are rather unsatisfactory Series AA and BB grade norms for 9th and 10th grades and high-school seniors are based on guidance center clients in the first two instances and on students applying for admission to the arts and engineering colleges of New York University in the last, certainly not a typical group of high school seniors

since it omits the 80 to 90 percent who do not go to college The college freshmen were students at New York University, the freshmen engineers were at New York University and Northeastern University, no significant difference having been found between engineers in the two institutions We have already seen, in the chapter on intelligence tests, what great differences exist between schools and between colleges, these norms are of local value, then, but can be no more than a rough guide to counselors or admissions officers in other communities and institutions

Series MA and MB norms are considerably better, the 9th grade norms representing three large cities, and the 12th grade norms, 60 New England schools

Stephens (756) administered the Revised Minnesota Paper Form Board to 2936 seniors and 3332 juniors, male and female, in all curricula in New England high schools, publishing norms based on them As he points out, these are higher than the old national norms, which we have seen to be strictly local The 1948 manual includes these norms, expanded by additional cases from subsequent samples

Hanman (329) tested 785 men in the educational program of the WPA in California, ranging in age from 20 to 65 with a modal age of 40 Their education varied as greatly, from none to the doctorate, with mode at 8 to 9 years The author concluded from his data that the old so-called national norms were too high (they were then based on 76 cases), failing, apparently, to take into account the fact that his was a selected, although large, sample, heavily weighted toward the lower end of the scale of education and ability Such heterogeneous and skewed norms have the values and uses of neither homogeneous and skewed nor of heterogeneous and representative norms

The sample studied by Baldwin and Smith (38) consisted of 975 women employed by the Eastman Kodak Co The group was divided into 16 to 25-year olds and 26 to 60 year-olds, norms for the younger group being somewhat higher than the original norms and those for the older group being somewhat lower Although this is in no sense a cross-section of adult women, and the norms are not general adult norms, they are useful in that they depict a large occupational population of varying skills The jobs to which they were assigned included unskilled repetitive jobs such as lens wrapping and highly skilled precision jobs such as final assembly and inspection of optical and mechanical equipment The 1948 manual includes these and other local but useful industrial norms, each set of which needs to be carefully studied by users

Standardization and Initial Validation. The completion form of the Minnesota Paper Form Board was one of the best tests in the Minnesota Mechanical Ability Battery, it had a correlation of 63 with its apparatus counterpart, and a validity coefficient of 52 against ratings of quality of shop work In revising the test and making it more objective the multiple-choice form was used, practice problems were included to insure understanding of the task, stencil scoring was utilized, and three time limits were tried out, the intermediate limit proving to be the best The test went through two revisions, and was standardized on college students High-school norms were then added It yielded a correlation of 40 with the Otis S A Test based on college students, and a correlation of 75 with scores on the original completion form Validity of the test was assumed to be demonstrated by the correlation with the original form and by the validity of that form, it was also ascertained by correlations of 49 with the mechanical drawing grades of engineering students and 32 with grades in descriptive geometry (617)

Reliability The uncorrected reliability based on the intercorrelation of the two revised forms of the test was found to be 79, while the split-half reliability was, corrected, 92 (617) This latter figure is made spuriously high, however, by the speeded nature of the test The retest reliability after periods of one or more years had elapsed was ascertained by Ebert and Simmons (233) with children aged 10 to 14, the age groups varying in number from 73 to 210 For 10-year-old children retested at age 11 the reliability coefficient was 87, at age 12 86, for 12-year-olds tested again at ages 13 and 14 the reliabilities were 87 and 80 It can safely be assumed, then, that the reliability is actually in the 80s and sufficiently high for individual diagnosis

Validity A criticism of time-limit tests such as this which is occasionally made by examinees or observers is that the imposition of a time limit makes the test a measure of speed and prevents it from measuring adequately the trait which it is designed to measure We have already seen that Baxter demonstrated the independence of speed (the time required to attempt every item once) and level (the number of items correctly answered in unlimited time), in the Otis intelligence test (p 108) Tinker (847) studied the roles of speed and level in the revised Minnesota Paper Form Board, confirming the finding that they vary independently Scores obtained in a standard time limit were found to consist primarily of speed, with level of difficulty at which the subject could work playing a lesser part Apparently tests would generally be

improved if they were administered as level tests, making possible the more nearly pure measurement of the trait being assessed, but the mixed speed and level scores now obtained for most tests are useful despite their impurity

Intelligence having been measured by tests which included spatial judgment items, one of the first steps in validating the Minnesota Paper Form Board has been to correlate its scores with scores on tests of general mental ability Sartain used two groups, one consisting of 46 inspectors in an aircraft factory (669) and the other 40 foremen also employed in an aircraft factory (671) Both groups took the revised Paper Form Board and the Otis S A Test, the correlations being 62 and 39, the reasons for the great difference are not clear, although the foremen may be a more homogeneous group The writer's intercorrelation of tests administered to 100 NYA youth yielded a relationship of 43 for the same two tests, which agrees not only with Sartain's foreman data but also with the relationship reported by Quasha and Likert (617) The NYA group was rather heterogeneous

The American Council on Education Psychological Examination was correlated with Paper Form Board scores in a study by Traxler (863), with 230 Merchant Marine Cadets as subjects The correlation of total scores was 42, for the linguistic scores it was 34 and for the quantitative it was 41 Bryan (123) tested art-school freshmen, correlating A C E part scores and Paper Form Board scores, for the spatial subtest of the A C E the correlation was 55 Army Alpha has been found (540) to have intercorrelations with the Revised Minnesota Paper Form Board which ranged from 35 and 31 at ages 14 and 15 ($N = 159$ and 109) but fell, unaccountably, at ages 13 and 16 to only 11 and 17 ($N = 86$ and 35) When the Revised Paper Form Board was correlated with the parent test (Geometrical Construction) of Army Beta (556) the coefficient was found to be 57, considerably lower than that of 75 between the original and revised forms of the Minnesota Paper Form Board referred to earlier The subjects were 9th grade boys in the Army Beta study, but college students in that of the two forms, which suggests that the larger correlation may have been obtained with the more homogenous group If this is so, then the revised test is more like the original Minnesota test than like the part of Army Beta from which they both originated

Manual dexterity is not an aptitude which one would expect to find playing a part in a spatial test as abstract as this is, but two studies have provided evidence concerning the degree of relationship Thompson

(824) found no relationship between either the Finger or the Tweezer Dexterity Test and the Revised Paper Form Board (− 08 and − 15), the writer obtained a correlation of 23 with the Minnesota Manual Dexterity (Placing) Test The true relationship is presumably about zero.

Mechanical comprehension has been seen, in the preceding chapter, to include spatial visualization among its components For this reason there is little to be gained here by repeating the data concerning the relationship as shown in various studies It should suffice to summarize by stating that the correlation with the O'Rourke is generally found to be about 40, with the Bennett about 35, with the Minnesota Mechanical Assembly Test about 48 (one study only), and with the MacQuarrie about 35 This means that a test of so-called mechanical aptitude may contribute materially to the prediction of success even when a good measure of spatial relations is used, for the score on the latter only partly accounts for the score on the former

Spatial Visualization as measured by apparatus tests such as the Minnesota Spatial Relations Test, the Crawford Spatial Relations Test, and the Wiggly Block should be correlated with the same ability as measured by the Paper Form Board, in order that the instruments and the trait may be better understood The writer found the correlation for the Minnesota test to be 59, his subjects being 100 NYA youths Jacobsen (396) found that for the Crawford to be 20 based on data from 90 mechanic learners Estes (240) reported that for the Crawford as 26 and that for the Wiggly Block as 31, with data obtained from 76 engineering freshmen Jacobsen's study, it will be remembered, reported a number of deviant results, disregarding it, therefore, we find only a moderate agreement among these rather different-appearing tests of spatial relations The fact that the Paper Form Board is more heavily saturated with general intelligence or inductive reasoning than the apparatus test explains at least a part of the failure to agree more closely It is also possible that there are differences between two- and three-dimensional spatial judgment, as the Crawford and Wiggly Block attempt to measure it, and if is true that when a test is as unreliable as the Wiggly Block it cannot often yield significant correlations with anything

Interest in mechanical and scientific activities as measured by Kuder's Preference Record was correlated with Paper Form Board Scores by Sartain (671), who found it to be negligible (r = 13 and 19) As the group consisted of foremen in an aircraft plant, who might be assumed to be homogeneous as to mechanical and scientific interests (high on the

former but low on the latter), this probably does not tell us much concerning the relationship between technical aptitude and interest

Three factor analysis studies involving the Revised Minnesota Paper Form Board threw further light on the subject of the traits measured by this test Morris (542) analyzed the intercorrelations of scores made by 56 9-year-olds to whom the Pintner-Paterson Scale of Performance Tests, the Porteus mazes, and Henmon-Nelson intelligence test, and others were administered, together with the Paper Form Board He found three group factors, which he called spatial relations, perceptual ability, and ability to discover patterns or a rule of procedure (induction) These resemble those found in the studies of the Minnesota Spatial Relations Test Murphy (556) used the Paper Form Board together with the Terman Group Test of Mental Ability, the Revised Army Beta, the Detroit Mechanical Aptitude Test, the MacQuarrie, and others, testing 143 9th grade boys Three factors emerged from this analysis mental manipulation of relations expressed symbolically (presumably induction), speed of hand and eye co-ordination (in the MacQuarrie particularly), and mental manipulation of spatial relations (in the Paper Form Board, parts of the MacQuarrie and Detroit, and part of Army Beta) Estes (240) gave the Paper Form Board, Crawford Spatial Relations, Wiggly Block, and A C E Psychological Examination (L and Q scores) to 76 engineering freshmen A factor analysis revealed one common factor, but this may be due at least partly to the small number of tests The implication, if correct, is that two- and three-dimensional tests of spatial judgment measure the same spatial factor, although imperfectly because of the different media Until further evidence is available, it seems legitimate to conclude that the Revised Minnesota Paper Form Board measures spatial relations, perceptual ability, and inductive reasoning, in that order, and that although it measures spatial judgment by means of two-dimensional media this ability is the same as that measured by three dimensional means

Grades and ratings of promise in training have been used as criteria in a dozen studies with this test Stanton (749) administered the original form to deaf boys and girls and obtained a correlation of .50 between scores and ratings of shop performance Jacobsen (396) used it with between 80 and 90 mechanic learners in a war industry, found that it correlated between .18 and .22 with fitness ratings, but the probable errors were so large as to make the relationships insignificant Ross (651) administered the Paper Form Board to 41 machine-tool trainees, but

published no correlations Apprentice pressmen were studied by Hall (325), with ratings of skill as a criterion, the correlation was 58 An attempt was made to differentiate between "good" and "poor" classes in an industrial and technical high school by means of the Paper Form Board, but Morgan (541) reported failure to discriminate, his subjects were 319 8th grade boys applying for admission to a technical high school

Several studies have used engineering students as subjects Berdie (78) obtained a low but significant correlation (22) between test scores and honor point ratios of 154 engineering students At the University of Maine, Brush (122) studied a group of more than 100 students, obtained correlations of 42 and 175 with first-year and 43 and 21 with four-year grades Physics grades at the University of Iowa were found to have a correlation of 26 by Stuit and Lapp (788) It can be concluded that the Revised Minnesota Paper Form Board does have value in selecting students for or guiding them in the consideration of engineering training, Brush found it one of the best aptitude, as contrasted with achievement, tests in his extensive battery, and it found a place in some of his best regression equations

Dental students were tested by Thompson (824), correlations with combined grades and ratings of 35 freshmen and 40 seniors being respectively 24 and 61, the difference is surprising, even when allowance is made for the fact that more professional work is included in the senior than in the freshman year

Art students have been studied with the Revised Minnesota Paper Form Board, on the assumption that spatial judgment is important in layout and related work Barrett (45) found that 40 art majors at Hunter College were significantly superior to 40 control students in spatial judgment, although the actual difference in scores was small Thompson (824) obtained a correlation of only 18 between the test scores and point-hour ratio for 50 fine-art students Bryan (123) used art grades as a criterion reporting a validity of 19

Success on the job has been studied more frequently with this test than with its apparatus counterpart, thanks to its group procedure Aircraft factory workers were studied by Sartain and Shuman in studies already described The former tested 46 inspectors and 40 foremen (669,671), the latter 263 engine and propeller workers, both skilled and semi-skilled (717), and 297 supervisors of several grades (716), ratings were the criterion in all instances Validity for Sartain's inspectors was 47, for his

foremen only .10 (as high as any in this study) For Shuman's workers it ranged from 16 to 59, depending upon the job moderately high correlations (.38 or above) were found for inspectors, machine operators, foremen, job setters, and toolmaker apprentices, the only low coefficient was for engine testers, for whom the Bennett Mechanical Comprehension Test had equally low validity, and for whom the critical scores on both tests were low, which suggests that the job may have been more clerical than mechanical In Shuman's other study the validity of the Paper Form Board for supervisors was found to be .33 The test would have improved selection by approximately 15 percent in each of Shuman's studies

Inspector-packers in a pharmaceutical concern were subjects of a study by Ghiselli (286), already described Ratings served as a criterion of the success of the 26 girls, for whom the Paper Form Board had a validity of .57 Stead and Shartle (750) report a correlation of only −.01 between scores on this test and ratings of 41 inspector-wrappers, but as they do not describe the job it is impossible to determine whether or not this finding is in conflict with Ghiselli's For can packers and merchandise packers they found validities of .28 and .48, for two groups of power-sewing-machine operators .31 and .48, and for put-in-coil girls the astonishing figure of −.52 This last group made the highest mean score of any tested by the USES research program as reported by Stead and Shartle Perhaps they were an able group who, bored by their routine jobs, actually tended to produce less than the less able girls The criteria in the jobs mentioned were based on output, and the numbers of subjects ranged from 18 to 46 For lamp-shade sewers and pull-socket assemblers, also tested in this investigation, the validities approached zero

Occupational differences in spatial visualization as measured by the Revised Minnesota Paper Form Board are suggested by Barrett's study of Hunter College art majors (45) which showed slight but significant differences between these students and control students in other fields, and by that of the USES (750), which found that put-in-coil girls and lamp-shade sewers were high average when compared to clients of the Adult Guidance Bureau of New York, and that the other workers listed in the preceding paragraph clustered around the 35th percentile The norms in the manual indicate, rather more helpfully, that engineering freshmen, at least in New York University, tend to score about five points higher than liberal arts freshmen, and that upper classmen in engineering curricula score about four points higher than freshmen Barrett's art majors made an average score equal to that of the engineering upper

classmen (raw score = 47), her controls an average equal to that of the engineering freshmen (raw score = 48) rather than the liberal arts freshmen, but the Hunter students were all upper classmen More comprehensive and varied occupational norms are badly needed for this test

Satisfaction in a professional curriculum, if not in the occupation itself, has been studied with the Minnesota Paper Form Board Berdie (78) gave the revised form to 154 engineering students and obtained curriculum satisfaction data by means of a modification of Hoppock's Job Satisfaction Questionnaire The correlation between spatial visualization and curricular satisfaction was only 06 The study can probably not be considered definitive, because a curriculum is something abstract and, unfortunately, often somewhat unreal to the student, whereas a job is usually something rather tangible Engineering students in particular are likely to be critical of the academic, despite ability and interest in technical matters. A study of vocational or job satisfaction might therefore yield different results

Use of the Revised Minnesota Paper Form Board in Counseling and Selection Although the Minnesota Paper Form Board is found to have a moderately high correlation with tests of general intelligence, more refined analyses have demonstrated that it is primarily a test of spatial relations, a special aptitude or distinct factor, and that the test is also somewhat saturated with quantitative perception and inductive factors It is the presence of this last, combined with the fact that some intelligence tests include spatial items, which makes the test correlate significantly with general intelligence tests A spatial relations test may therefore make a distinct contribution to some test batteries

Maturation of ability to judge spatial relations seems to come in the early teens, with little if any increase after age 15 or 16 This suggests that adult occupational norms should be usable with high school juniors and seniors, and perhaps even with sophomores

Occupations for which the test has been found to have significance include professions such as engineering, art, and dentistry, skilled trades such as toolmaking, job setting and aircraft engine inspection, and semiskilled jobs such as inspection and packing of merchandise, cans, and other objects, power-sewing-machine operation, and electrical assembly Supervisors and foremen of both skilled and semiskilled workers also tend to make superior scores on this test

In schools and colleges the test should be found useful for counseling concerning the choice of trade courses, engineering curricula, dental

training, and the professional study of art Presence of the trait in a high degree cannot be considered a good prognosticator of success, because of the importance of other aptitudes and traits, but its relative absence in an individual can be considered a danger signal Despite the importance of spatial visualization in tests of so-called mechanical comprehension, the correlation between these two types of tests is low enough to prevent the use of both from being a duplication

In guidance and employment centers the use of the test can be comparable to that in educational institutions when choice of training is involved It can be of value also in selecting individuals who are likely to adapt quickly to the demands of assembly work and machine operations in new jobs in which they might be placed

In industrial personnel work the Minnesota Paper Form Board can be valuable in the selection of adaptable workers for semiskilled employment, for the evaluation of workers on the job whose skills may be most readily utilized in new assembly or machine operations, and also in the selection of apprentices for training in the skilled trades In any such selection or evaluation program other indices should also be obtained, and here too a good mechanical comprehension test, an intelligence test, oral trade tests, and evidence concerning leisure-time activities which throw light on aptitudes and interests may be important data

CHAPTER XII

AESTHETIC JUDGMENT AND ARTISTIC ABILITY

ARTISTIC ability has been broken down into six factors in studies conducted over the past twenty years by N C Meier and his students at the State University of Iowa (519) The analytic procedures used were partly biographical, partly mensural, and should not be confused with the more objective procedures of factor analysis, but in the absence of analyses utilizing completely objective methods Meier's conclusions after years of research provide the best available insights into the nature of artistic ability

The six factors listed by Meier include *manual skill,* as evidenced in studies of the family histories of artists, *energy output and perseveration,* revealed in studies of biographies, *aesthetic intelligence,* by which Meier means spatial and perceptual aptitude as measured by Thurstone's tests, *perceptual facility,* or the ability to observe and recall sensory experiences, which this writer cannot distinguish clearly from the perceptual ability just mentioned, evidenced in biographical material and in a test of recall of observed material after intervals of 10 days and of 6 months (212), *creative imagination,* defined as an ability to organize vivid sense impressions into an aesthetic product, trait concerning the existence of which no satisfactory evidence has been adduced, save the aesthetic product itself and the uniqueness of ink blot interpretations (212) which may actually indicate personality deviation, and *aesthetic judgment,* considered to be the most important single factor in artistic ability, defined as the ability to recognize unity of composition and believed by Meier to be not the application of a series of rules, but rather something which is innate in the neuro-physical constitution and modifiable by experience Some of these factors, such as aesthetic intelligence, are treated as complexes which can or will be broken down into underlying unitary traits of the Thurstone variety, others, like aesthetic judgment, are considered themselves basic and unitary

As in any occupation, success in art may be due to various combinations of the abilities and traits just described The artists whose lives Meier has studied are believed to have excelled in some of the abilities listed, although not necessarily in all Meier developed his Art Judgment Test first, because of his conviction of the primary importance of this aptitude, his plans for subsequent work, financed by the Spelman and Carnegie Foundations, called for the development of tests for the two other abilities in his list which are not presently mensurable, namely, perceptual facility and creative imagination The writer is unaware of any practical tests resulting from this work Of the three other traits, the manual and intellectual factors are currently well measured by existing tests, already described, while the emotional characteristics, as we shall see, have so far not lent themselves to satisfactory measurement

In appraising artistic promise it would therefore seem well to use, a), tests of intellectual ability, particularly those tapping spatial factors Tiebout and Meier (844) found that 50 outstanding artists selected from 5,500 listed in the Biography of American Artists had an average Otis IQ of 118, with their successes predominantly in the verbal and spatial items, b), tests of manual dexterity, although, as we have seen in that chapter, there is little in the way of normative material to assist one in test interpretation (presumably an average score or better would be desired), and, c), tests of aesthetic judgment, discussed in detail in this chapter Other data must be gathered by means of techniques other than tests These might include the expert appraisal of the counselee's sketches, paintings, or other art products, the summarization of experience in artistic avocations and activities, and the evaluation of motivation to persevere in art as shown in discussions of artistic activities and aspirations

Aesthetic Judgment

Aesthetic judgment emerges as the one trait in Meier's list of six which may be considered a candidate for discussion as a mensurable special aptitude not dealt with in this book under some other heading It is for this reason that it is singled out for treatment toward the end of our list of special aptitudes, and before batteries of aptitude tests and measures of personality and interest are taken up

There are two well-known tests of aesthetic judgment the *Meier Art Judgment Test* (a revision of the Meier-Seashore Art Judgment Test) and the *McAdory Art Test,* the original editions of which were both published

in 1929 The graphic material used by Meier was more or less timeless, for it included masterpieces of art which appear to be able to withstand the temporary shifts of fashions and of schools, that used by McAdory was more transitory, for it included textiles, clothing, furniture, and architecture, and it need hardly be pointed out that the dresses and automobiles of the late 1920's no longer seem to represent the acme of good taste in composition McAdory has until recently done no further work with her test (now being revised), while Meier has maintained his interest and his production The McAdory is for the time being of purely historic interest, a summary of work with it will be found in Kinter (425) Other similar tests are too new to have been studied The Meier alone will be dealt with in this book, as the only art judgment test of practical significance for the psychometrist or counselor Two tests of so-called creative artistic ability, both in reality worksamples, are also briefly treated here, for lack of a more appropriate place

The Meier Art Judgment Test (Bureau of Educational Research and Service, 1940 Revision)

The first edition of this test, published in 1929, was known as the Meier-Seashore Art Judgment Test It was revised and published as the Meier Art Judgment Test in 1940 During the intervening years Meier and his students at the State University of Iowa conducted a number of important studies in the nature of aptitude for artistic work, summarized in the 1941 Yearbook of the National Society for the Study of Education (520), in a brief monograph chapter (662), and in his broader treatise on *Art in Human Affairs* (521) Meier's perseverance in the study of artistic ability has given his institution a leading place in this field which has been rivalled only by the leadership in the study of musical aptitudes which it exercised under Carl Seashore, it is interesting to note that a mid-western state university has led in the "impractical" field of aesthetic research.

Applicability The revised form of the Meier Art Judgment Test, like its predecessor, has been standardized on junior and senior high school students and on college students Greene (309 395) points out that the grade norms for the Meier-Seashore Test show nearly chance success at the 8th grade level, and refers to other studies which showed that the ranking of pictures by 10-year-olds was similar to that of average adults, that of 7-year-olds already showing considerable agreement As the latter studies did not use the same types of materials as the Meier tests, it is

not possible to draw precise conclusions from a comparison, it seems probable, however, that the judgments required by the Meier tests are more refined than those involved in the other studies, and that this more refined type of aesthetic judgment matures later The median score for junior high school students on the revised form is 88, whereas that for senior high school students is 99 This difference is presumably due in part to selection, but, as the test has a very low correlation with intelligence, it may be concluded that it is due primarily to developmental differences Meier seems to attribute this largely to experience in his book (521 131) but not in the manual (pp 15–16) Apparently aesthetic judgment is still developing during the middle teens, making age norms necessary As in the case of so many other tests, tables making possible the conversion of age-group percentiles into occupational percentiles would be highly desirable It is noteworthy that training in art has been found to have little effect on scores (139)

Content The Meier Art Judgment Test, 1940 revision, consists of 100 pairs of pictures, printed in booklets with one pair per page on one side of the sheet only The pictures are largely paintings, sketches, etc, which are generally recognized as works of permanent merit, others are vases and other *objets-d'art*, all were included because of agreement concerning their merit by a group of established artists and because of high biserial r s with total scores In each pair, one member is the unaltered reproduction of the original work, while the other member is a slightly modified version The modifications are designed to make the composition, form, etc, less pleasing to the eye, the nature of the difference is pointed out to the examinee The examinee's task is to decide which picture he prefers in each pair, with no knowledge of which is the original picture (the paintings are not so well known that subjects are likely to recognize the original)

Administration and Scoring The Meier test can be administered either individually or in groups, but, as there is no time limit and there is great variation in the amount of time required to complete it, it is not a convenient test for group administration at any time other than the end of a test battery It is usually completed in less than one hour Scoring is by means of a stencil, is simple and objective

Norms The 1942 manual for the revised test provides norms for 1445 junior high school, 892 senior high school and 982 college art school students The students were "interested in art" "for the most part," making the norms representative of neither general population nor art

students, except at the college level The 25 schools represented were scattered throughout the whole United States

Standardization and Initial Validation In the initial standardization work for the earlier form of the test nearly 600 pairs of items were tried out on over 2000 pupils in various types of schools and colleges The 125 which were then retained were those which had the most discriminating value and those which were most favored by a group of experts The current revision includes 100 items selected in a similar way during the eleven years intervening between the two forms of the test The method of selecting the items may perhaps be considered evidence of validity, for the answers scored "right" are those which are chosen by high scorers on the test as a whole, and are those which are chosen by established artists That no established artists made low scores, and that some untrained persons made high scores, was taken by Meier as an indication that the test measured an aptitude rather than the effects of specific training (522), although he has since modified his point of view to allow a somewhat more important role for experience (521) Correlations with intelligence test (Terman Group, Standard Binet, Thorndike) scores were found to range from − 14 to 28, indicating that it was not a measure of general intelligence Comparable data for the new form have not been published

Reliability The earlier form of the test had retest reliabilities which ranged from 61 to 65 for non-art students, and from 69 to 85 for art students (518, Leighton cited by 425, 141) These are lower than is desirable in a test used in individual diagnosis, making caution necessary in its use The reliability of the 1930 revision ranges from 70 to 84, those two lowest being based on students of Pratt Institute and a junior high school, the two highest in an art school and a senior high school (grades not specified) It is to be regretted that they were not raised for more accurate diagnosis, but as Meier points out the test is really only a screening device, which makes the reliability adequate

Validity All but a few of the published studies of the Meier Art Judgment Tests are based on the older edition Because of the similarity of the two revisions they are briefly discussed here, together with the little new material available

Items in the early form were analyzed by Brigham and Findley, reported by Kinter (425), who calculated the biserial coefficient of correlation between items and total score, the correlations ranged from − 02 to 53 Perhaps this partly explains the relative unreliability of the first edition of the test, which clearly contained dead wood The revision used

Brigham and Findley's data (522 14) to select the 100 best items, thereby correcting this defect When old records were checked with the new key, greater differentiation was found

Intelligence test scores, correlated with scores made on the original form by art students (141) and by college students (248), were only slightly if at all related to aesthetic judgment (28 and 03) These findings agree substantially with Meier's

Spatial visualization test scores might logically be expected to be related to art judgment scores, since the aesthetic judgments involve the arrangement of objects in space Brigham and Findley (in Kinter, 425) found a correlation of 37 with the College Entrance Examination Board spatial test, showing that the two aptitudes do have something in common Unfortunately no other intercorrelations of such tests have been located, although the data for their computation have been available (45) A factor analysis of a battery of tests of these two types, plus others of art information, perceptual ability, etc , might throw considerable light on the nature of art judgment

Artistic judgment as measured by the McAdory test correlated only 37 (907) and 27 (139) with the Meier-Seashore Test, a difficult finding to explain The Lewerenz Test of Fundamental Abilities in Visual Art is related only to the extent of 53, but this is not surprising in an *ability* test it is, in fact, rather gratifying as an index of validity

Art grades were related to scores on the first edition by Brigham and Findley, who found the surprisingly high validity of 46 for a group of 50 students at Cooper Union but concluded, according to Kinter (425 61), that the test did not have sufficient discriminating value—perhaps because of the inclusion of poor items No data are available for the new form, for which they should be at least as good

Ratings of creative artistic ability have been somewhat more extensively used as a criterion of the validity of the Art Judgment Test Carroll (139) found a correlation of 40 between these two variables, Morrow (543) found a validity of 48, and cited one by Jones of 69 Apparently the test has considerable value in selecting the students who manifest promise in their art courses

The differentiation of occupational groups by means of the Art Judgment Test has been demonstrated, primarily with students but to a lesser extent with artists and art teachers The manual for the first edition shows that art teachers made higher scores than art students or students in general, but no critical ratios were computed Eurich and Carroll

(242) found that art majors ranked 8 14 points higher than other college students on the old form, which seems especially important in view of their finding that training had no effect on scores, the difference was statistically significant, the groups large Barrett (45) confirmed this finding with the revised test, art majors at Hunter College scoring six points higher than non-majors on the average, a difference which was significant at the one percent level More helpful than any of these data would be correlations between pre-training scores and success in art work, but no such data are available

Vocational satisfaction has not, apparently been related to art judgment in any studies

Use of the Meier Art Judgment Test in Counseling and Selection The evidence concerning the Meier Art Judgment Test indicates that it measures an ability which varies from person to person, is found in a higher degree among artists than among non-artists, is possessed by some untrained persons in a very high degree, is distinct from intelligence and only moderately related to spatial visualization, is not much influenced by training in late adolescence, and is related to success in art training This ability therefore seems to be an aptitude in the narrower sense of the term

Development of aesthetic judgment appears to continue well into adolescence, making age norms desirable Just when development begins to level off is not clear, however, as the ability is a relatively complex one, it may be safe to assume that levelling off takes place in the late teens or early twenties Carroll's work suggests that devolpment is more a matter of maturation, at least in late adolescence, but this question needs further investigation

Occupations in which aesthetic judgment may be important have unfortunately not been extensively investigated, Meier s efforts having been absorbed in the study of other problems That artists excel in it has been demonstrated, but the writer knows of no data which show the role which it plays in other fields, such as clothing design, dramatic production, architecture, and landscape gardening

In schools and colleges the Art Judgment Test should be useful as a means of locating students who may have special talent and deserve special opportunities for artistic training, special attention in art courses, and encouragement to capitalize on extra-curricular opportunities for the development of their talent, whether for vocational or for avocational purposes It can also be useful as a selection instrument in art schools,

although at this stage, as in counseling to a lesser extent, the evaluation of artistic production often yields more helpful information In judging a client's or applicant's art work it is necessary that the judge be not only an artist, but an artist who is used to appraising the work of beginners in the light of the amount of training they have already had When, for special reasons, samples of a counselee's work are not available, it may be desirable to administer a worksample test of artistic ability such as the Lewerenz Tests in the Fundamental Abilities of Visual Art or the Knauber Art Ability Test, both of which were designed to measure creative ability in art (described below)

In guidance centers the use of the test is similar to that in schools and colleges, whether for counseling purposes or for selection in connection with training programs It has little place in the evaluation of employment applicants, as these are normally already trained in art and can better be judged by their work, unless an especially important position is to be filled and it is desired to have a comprehensive study of the applicant

The business and industrial use of the Art Judgment Test is extremely limited, for reasons just given It may, however, prove quite valuable at times when nonartistically trained personnel are to be selected or transferred to work in which ability to judge good form and composition are important, for example, in certain retail trade jobs of a merchandising type

Creative Artistic Ability

As was mentioned earlier in this chapter, tests of so-called creative artistic ability are in reality worksamples devised to measure the subject's ability to construct a good artistic design or to utilize the concepts, vocabulary, and tools of the artist As such they hardly belong in a discussion of aptitudes in the narrower sense of the term, but logically should be taken up in connection with custom-built tests or, if there were enough such to warrant such a classification, with worksamples In this case it seems more practical, however, to treat these tests in the chapter dealing with another special aptitude, the importance of which is seemingly limited to the same occupations If Meier makes available the promised battery of art tests, a change in the organization and location of this material will be warranted, giving it a section in the chapter on custom-built batteries of tests

The two worksample tests dealt with here are the Lewerenz and the

Knauber Because of the similarity of content and the lack of subsequent studies of their validity, both are briefly discussed, the Lewerenz being given more space as it is a more manageable test

The Lewerenz Tests in the Fundamental Abilities of Visual Art (California Test Bureau, 1927)

Applicability The test was designed as a measure of creative artistic ability, for use in school systems It was standardized on children in grades 3 through 12 It can also be used with young adults who have had no further artistic training

Contents, Administration, and Scoring Because of the independence of the separate parts of the test, they are best described in detail individually

Test 1 Fifteen sets of drawings with four pictures to a set (multiple-choice), including bowls, friezes, cornices, etc , varying from good to bad in proportion and balance Two parts, recognition of proportion in standard forms, and problems of abstract proportion and balance Time, 10 minutes Score equals number right

Test 2 Ten sets of dots in varying numbers The subject is told to draw any subject he chooses, using all the dots in each space with straight or curved lines, then to write one word in the space to indicate what he has drawn The arrangement of dots varies, to permit formal and fanciful interpretations Time, 20 minutes Score is obtained by comparing drawings with six graded rating sheets

Test 3 Ten drawings ranging from simple to complex The subject is required to indicate omissions of shades and shadows, the light being considered as coming from the left Time, 5 minutes Score is the number right

Test 4 A vocabulary test, utilizing the matching method in five ten-word sections dealing with materials, craft processes, graphic processes, drawing terms, and pictures Time, 20 minutes Score is the number right

Test 5 A black vase form mounted on a white background is exposed to the subject for two minutes After it is removed the subject is instructed to draw it from memory, on a test blank which shows the top and bottom of the vase with a vertical line through the center Time, 5 minutes Scoring is by a stencil

Tests 6, 7, and 8 deal with ability to analyze problems in perspective cylindrical, parallel, and angular The subject may use a ruler in correct-

ing incorrectly drawn lines in each of the three tests Time, 5 minutes for each test Score is the number of correct responses

Test 9 A color chart with six known colors at the top, below are 46 "unknown" variations divided into four sections The initial letter of the known six colors is used to indicate the one predominant known color in each of the unknowns, by means of a six-response type multiple-choice technique Time, 20 minutes Score is the number correct

Norms Norms are available for elementary grades, junior high school, and senior high school, based on an unselected group of 1100 pupils Separate norms may be used for part scores

Standardization and Initial Validation As has just been stated, the tests were standardized on a supposedly typical group of school children Various comparisons were made by the test author with art students, the correlation with art grades being 40 and the rank correlation between performance and predicted ability 63 In subsequent studies, summarized by Kinter, Lewerenz found a correlation between his tests and a test of intelligence of 155 for a group of over 1000 children Sex differences were also reported, girls being superior to boys in all but originality and ability to analyze

Reliability A retesting of 100 pupils in grades 3 to 9 after an interval of one month yielded a reliability coefficient of 87 (manual) No other such data have been located

Validity Few studies have been made involving the Lewerenz tests by persons other than the test author, which is to be regretted in view of the fact that it is the more manageable of the two well-known tests designed to measure creative artistic ability Wallis (907) correlated the test with the Meier Seashore and McAdory tests, finding correlations of 53 and 58, higher than that between the last two, which are supposedly more similar (37)

Use of the Lewerenz Tests in Counseling and Selection From the above material it is clear that the Lewerenz tests are measuring, with considerable reliability, various factors which are rather distinct from intelligence, which have a substantial relationship with achievement in art, and which vary with age and sex An analysis of the content suggests that these factors have to do with visual and creative artistic abilities, but too few relationships have been determined, and no factor analyses have been carried out, to enable one to draw adequate conclusions On the basis of the available evidence, however, one may tentatively conclude

that the tests have practical value in selecting students with sufficient promise for further training in art

Because comparatively little is as yet known about it, scores on this test must clearly be supplemented by a variety of other information, such as Meier scores, data on art training and interests, intelligence, ratings of art work, etc

The Knauber Art Ability Test (Distributor Psychological Corporation, 1927, revised 1935)

Applicable at or above the junior high school level *Contents* are drawings in which the subject creates or completes drawings or locates errors, they yield seven measures of presumed components of art, such as long and short-time memory, observation, accuracy, creative imagination, ability to visualize and to analyze, etc *Administration* involves no time limit, but the test normally takes about three hours Drawings are rated on a three-point scale *Norms* are based on 1366 students from 7th grade to university sophomore, are in terms of grade percentiles *Standardization and validation* are described in the manual and in an article by Knauber (439) The present form is standardized on 1366 cases, after trials of other forms on 300 art students and 550 art students Art students make a median score of 95 compared to that of 52 for non-art students, art teachers make a median score of 123 contrasted with 61 for other teachers With 42 art students as subjects, the correlation with the Meier-Seashore Test was 57, with the Lewerenz, which it should resemble more closely on a priori grounds, 64 *Reliability* Retest reliability after one year was 96 (438), by the split-half method it was 95 *Use in counseling and selecting* seems justified by the fact that the test distinguishes between the various levels of artistic ability as shown in the group differences reported No data are available, however, on the effects of training, which might account for these differences, except the evidence showing high reliability over a period of one year The test does appear to measure creative ability, if the nature of the items may be taken as evidence of validity However, in view of the present limited knowledge of the test, scores must be used with considerable caution Those making high scores may, if other evidence such as art judgment tests, intelligence, interests, and ratings of art work, is favorable, be encouraged to continue training in art, cases making low scores should be investigated further before recommendations are made

CHAPTER XIII

MUSICAL TALENTS

TO TURN our attention to musical aptitudes in this chapter, as to artistic in the preceding chapter, is to risk abandoning the logic of the organization of the book as a whole For in this text the focus is first on psychological characteristics, whether they be aptitudes, skills, or traits, then on the means of measuring them, and finally on the vocational and educational significance of the ability or trait being measured The use of the terms "artistic' and "musical" implies an orientation which is primarily occupational Useful as this latter approach is when judging a person's fitness for a specific occupational field or when devising or selecting a battery of tests for a single area, it is not, on the whole, as helpful as the psychological approach is to the counselor who seeks an understanding of the person with whom he is working and who hopes, through a sharing of that understanding with the client, to help him to make appropriate vocational plans In this chapter as in the preceding, however, the focus on the occupational field is brief and introductory to the discussion of specific aptitudes which happen to be important primarily to one family of occupations The aptitudes, in this instance, are physical capacities which have been found to be fundamental to success in music, they include such abilities as sense of pitch, sense of rhythm, and sense of time They are treated in some detail below, in connection with the Seashore Measures of Musical Talents

Music being a creative aesthetic occupation, it seems likely that many of the traits which have been shown or are presumed to be of importance to success in artistic occupations would also play a part in musical success Seashore has studied these in an early monograph (690) and discussed them in his more recent general treatise of the psychology of music (693), and the list does indeed tend to parallel that of his colleague Meier in the field of art Manual skill is considered necessary for instrumental work in music, as for the use of tools in art, energy output and perseverance is deemed important in music too, with its requirement of

hour after hour of routine practice, creative imagination is presumed to play a part, not only in the composition of new works but also in the interpretation of existing works, and emotional sensitivity may be thought to be important in both creative and interpretive work, if the musician or the artist is effectively to portray feeling and to play upon the emotions of others Intelligence may be assumed to be increasingly important at the higher levels of musical endeavor, while it may not be important in a blues singer, Stanton's studies at the Eastman School of Music (748) showed that intelligence is important in mastering the more abstract aspects of music And, finally, Seashore's investigations (662), confirmed by those of Stanton and others, have shown that the physical capacities measured by his tests are basic to musical success

As the preceding paragraph implies, the only factors presumed to be important to success in music which have satisfactorily been demonstrated to be related to achievement in that field are intelligence and Seashore's psychophysical capacities The writer has seen no investigations other than the tentative early study by Seashore (690) which demonstrated that musicians are superior to the general population in manual skill, energy output, or creative imagination, or that scores on measures of these factors are correlated with musical success There is some evidence which suggests that musicians may be more sensitive emotionally than the general population, for the writer (791) found that male amateur musicians who played in symphony orchestras were significantly more likely to be unmarried, dissatisfied with their social life, and dissatisfied with their occupations than were other men of the same age and socio-economic status If maladjustment is a sign of emotional sensitivity, then the hypothesis is perhaps validated, but it is possible that there is such a thing as emotional sensitivity without maladjustment, and that it is sensitive persons who are not maladjusted who make the best musicians In any case, the writer's subjects were amateur, not professional, musicians It cannot therefore be said that it has been demonstrated that emotional sensitivity plays a part in success in music

In view of the demonstrated importance of Seashore's physical capacities in musical success, the infrequency with which they play a part in other fields, the lack of evidence concerning the significance of other abilities in music, and the general rather than specifically musical nature of the other characteristics which are presumed to affect success in music, it seems legitimate to discuss Seashore's tests and the capacities which they measure under the heading of musical aptitudes or talents Other

similar tests, described by Greene (309 425–438), are not dealt with here because they have not been so thoroughly studied

The Seashore Measures of Musical Talents (RCA Manufacturing Co, 1939, since 1949, Psychological Corporation)

The initial work on the measurement of physical capacities which might be important to success in music was begun by Seashore before World War I As in the case of other psychologists who were then developing new measuring instruments, he continued his work during the war, applying it successfully to the selection of submarine detection men in the Navy The first edition of the test for general use in musical guidance and selection was published soon afterwards, in 1919 As a pioneer in the study of the psychology of music, and aware, apparently, of the value of focusing his research energies on one promising field, Seashore continued to work with his tests, attracted graduate students who carried out additional studies, and found financial support to press his and his students' investigations As a result, his laboratory at the State University of Iowa became the most active center for research in the psychology of music and in the prediction of musical success in the United States, and his tests are, together with the Stanford-Binet and Strong's Vocational Interest Blank, among the best known, most widely used, and most thoroughly understood instruments in the field of psychological measurement The tests were revised and a second edition published in 1939 (662)

For these reasons, the tests are treated here in some detail, even though the frequency of their use in counseling is somewhat limited because of the relatively few persons in musical occupations Were it not for this fact, they would be dealt with at much greater length, as an illustration of the thorough type of work and multiple approaches which are needed in making vocational tests useful

Applicability The first edition of the Seashore tests was designed for use at any grade level, from the first grade to adulthood Because of the effects of motivation and attention on the test scores, however, the revised manual recommends that the tests be used beginning with the fifth grade, that is, with children of about ten years old This is acceptable to Seashore as a minimal age because it is also early enough to make possible serious planning for musical training if it seems warranted

The norms for the revised tests indicate that scores tend to increase somewhat with age, for there is a steady increase in the means from

grades 5 and 6 to adulthood Although these differences are slight, amounting to only one or two points, they might conceivably be interpreted as showing that the abilities in question are still maturing The ranges of scores are the same, however, at the different age levels, and the reliabilities are somewhat higher in adulthood than in adolescence (median r = 82 in adulthood, 78 in adolescence), facts which suggest the validity of Seashore's contention that the lower means of younger people are due to problems of concentration, attention, and similar administrative factors If this is so, it becomes important to take especial pains to establish good rapport when testing school-age children and to test in two or three sessions. Seashore (690) and Stanton (748) have shown that training and experience, e g , three years in a school of music, do not influence scores The tests are therefore as applicable to adults as to children, and vice versa

Content The tests consist of two series of three double-faced twelve-inch phonograph records each Series A is made up of wide range tests suitable for survey or screening purposes with heterogeneous groups, while Series B has a higher base and "ceiling" in order to make it more diagnostic at the higher ability levels and with music students The six capacities measured by either series are Pitch, Loudness (formerly called Intensity), Time, Timbre, Rhythm and Tonal Memory The 1919 edition contained a test of Consonance, for which Timbre was substituted No verbal description can convey an adequate idea of the specific content, but it may help those who do not have access to the tests to describe the Pitch Test, for purposes of illustration, as a series of pairs of musical notes One member of each pair of notes is higher than the other, sometimes the higher note comes first, sometimes last, in the pair, in later pairs the two notes are of more nearly the same pitch than in the first, the notes becoming more and more alike in pitch as the test progresses As a result, a point is reached at which it is virtually impossible to decide which note is higher This point comes early in the test for those lacking in pitch discrimination, late in the test for those who excel in it The other five tests are built on similar principles

Administration and Scoring The manual for the 1939 edition gives quite adequate directions for administering the tests, which require about one hour Several points deserve special emphasis, however, because of the unusual nature of the medium The records used must be in good condition, neither scratched nor warped So also should be the record player, adjusted to play loud enough to be heard throughout the

room, and at the standard speed of 78 r p m As the records are monotonous, capturing the interest and retaining the co-operation of the subjects is especially important in a paced test such as this a little wandering of the attention can spoil a test score The manual recommends that examinees lean slightly forward in a poised position which facilitates concentration Most unusual in the testing procedure is the desirability of demonstrating the tests by playing parts of each record before testing, the examiner gives the directions, then plays a few items near the beginning of the record, asking all examinees to respond orally, and permitting time for questions He plays a few more items nearer the end of the record, again asking for group responses and allowing questions This is to familiarize all subjects with the unusual type of test item, and to make it truly a measure of capacity It might be objected that the test is spoiled by familiarization with the specific contents, but experimentation has shown that practice does not vitiate the test if the excerpts from the records are not consecutive (9, 246) Responses, in terms of "high, low," "strong, weak," or similar terms, are recorded on simple answer sheets which can be purchased or mimeographed, scoring is done by comparing responses with a key or a homemade stencil, and counting the number of correct answers The tests can be given more than once for the sake of greater reliability, and the scores averaged, which single fact more than any other brings out the fundamental difference between this and other aptitude tests!

Norms Decile norms are provided for 5th and 6th grade pupils, 7th and 8th graders, and adults, for Series A tests, and for adults only for the Series B tests No separate high school norms were deemed necessary, because of the small differences, already referred to, between 8th graders and adults The normative tables do not indicate the number of cases on which the standardization was based, but the table of reliabilities in the manual makes it clear that the numbers in each grade group varied from about 1000 to 1700 pupils, depending upon the test, and from 600 to 1100 adults, the smaller numbers of cases being for Series B There is no indication as to how the samples were selected, as Series A is designed as a survey test it should be a cross-section of school children and adults in general for that series, and, for Series B, the diagnostic test, a group of adults studying music The manual is defective in not making the nature of the samples explicit

Standardization and Initial Validation Adequately to describe the extensive and intensive standardization and validation studies carried out

with the Seashore music tests by Seashore, his students, and other psychologists interested in music would require far more space than abilities of such limited occupational significance merit in a text such as this In fact, even the full-sized volume in which Seashore discusses his twenty-five years of work with the tests is tantalizing to a scholar because of its generality and lack of specific data on what was done and with what results For present purposes, it seems best to survey a few of the studies of the validity of the tests, referring those interested in their standardization to the monographs by Seashore and his colleagues (690,692,662)

Reliability Farnsworth (246) reviewed the studies of the reliability of the old form of the tests in 1931, 88 in all, and concluded that only the tests of pitch and tonal memory were sufficiently reliable for use with individuals Drake (210), for example, found that the better tests had reliabilities of about 86, these were odd-even reliability coefficients, corrected by the Spearman Brown formula, and might be spuriously high in a test which is paced and therefore somewhat speeded However, Larson (453) retested children and adults with substantially the same results The revised battery has higher reliabilities, on the whole for Series A they range from 69 to 84 at grades 5 and 6, from 69 to 87 for 7th and 8th graders, and from 62 (the next higher is 74) to 88 for adults The median reliabilities at the same levels are 78, 785, and 82 For Series B the coefficients are somewhat lower 70 to 89, with a median of 735 Tonal memory is the most reliable test in the new battery, with pitch and loudness about equally good, while timbre, which replaced the unsatisfactory test of consonance, is the least reliable It seems surprising that what appear to be immutable physical capacities are measured with less reliability than some more strictly psychological factors, perhaps this is due to the large number of fine discriminations which must be made, and to vagaries of attention, rather than to the nature of the trait or defects in the tests

Validity Most studies of the validity of the Seashore tests have been concerned, as one might expect, with the relationship between scores and variables such as intelligence, music grades, and success as a musician In the revised manual and related publications (662), however, Seashore has taken a new and different position Although the validation studies have tended to demonstrate a considerable degree of predictive and occupational differentiating power, he now seems to feel that the validity of the tests lies in their accurate measurement of basic capacities which are utilized by musicians, rather than in the degree to which they are

correlated with success in musical training or performance The difference may seem a fine one, but it may be made clearer by explaining that in the latter approach one correlates test scores with grades or ratings, whereas in the former one analyzes the performance of musicians in order to ascertain to what extent they reveal high degrees of pitch discrimination, sense of timbre, etc To the writer, this seems like a reversal of the natural order of things, for surely one should analyze the job to ascertain what factors seem to be important in it, then construct tests to measure them, and then, as validation of both the job analysis and of the tests, correlate scores on the tests with criteria of success on the job If there is no relationship between the measures and success, it matters little what the analysis showed Perhaps Seashore did not intend to convey the impression that he had thus reversed his approach, or perhaps it was simply that, having found objective methods of analyzing the performances of musicians (692), his interest in the technique caused him to lose sight of its place in the prediction, as opposed to the analysis, of musical performance Be this as it may, there are a number of helpful studies of the predictive value of the musical aptitude tests in their older form, comparable studies of the essentially similar new form have yet to be published, research of this type having been interrupted during World War II and Seashore having been retired

Intercorrelations of the original six tests were reviewed by Farnsworth (246), who found them to have a median intercorrelation of 48 for college students and 25 for elementary and junior high school pupils This suggests that the capacities measured by these tests are not as completely independent and basic as Seashore believes them to be, suggestion apparently confirmed by Drake's factor analysis (211) of the five best Seashore tests, the Kwalwasser-Dykema tonal movement test, and two new tests, one of memory and one of retentivity, which revealed one common factor and three group factors underlying them It may be, for example, that senses of pitch and rhythm underlie tonal memory

Intelligence has repeatedly been found to have little relationship to Seashore scores Farnsworth's review (246) covered the earlier studies of this topic, sixteen in all, with a median correlation of 10, the range being — 08 to 45

Grades in music courses have less often been used as a criterion of success, perhaps because they have not seemed sufficiently representative of musical ability Larson's finding of a correlation of 59 between composite Seashore scores and grades in the first course in music theory at the East-

man School of Music seems rather high, a correlation of .31 between Seashore tests and grades in a college of music was reported by Highsmith (370), which seems more in line with probability Intelligence tests were found more useful in this latter study (r = .42), and were included in the Eastman School battery (748)

Ratings of musical ability have not yielded such satisfactory results Mursell (559) reviewed such studies and drew the conclusion that the tests were invalid In view of studies such as Stanton's (see below), which have utilized objective procedures and have demonstrated considerable validity in the tests, it hardly seems justifiable to make such drastic judgments on the basis of data as subjective as ratings Not only have ratings generally been proved unreliable (810), but in studies such as those in question the subjects rated were all sufficiently able in music to be active students, a select group, thereby narrowing the range of both ratings and scores and artifically attenuating the relationship In such circumstances the making of ratings is more difficult and the product therefore less reliable than ever

Completion of musical training seems a much more objective criterion of success than rating, even when the effects of financial factors are recognized Stanton (748) made a ten-year study of the Seashore tests at the Eastman School of Music in Rochester More than 2000 entering students were tested, and the test results were not used but simply filed until criterion data were available four years later An analysis was then made of the relationship between test scores and the completion of training in music The results of Seashore tests were combined with intelligence (Iowa Comprehension) test scores and teachers' ratings to provide a "cumulative key" or overall predictor It was found that 60 percent of those who were rated "safe" risks on this basis had graduated in the normal amount of time, 42 percent of those who were classified as reasonably good risks and 33 percent of the fair risks graduated, in contrast with 23 percent of the poor and 17 percent of the very poor risks The case histories of the high-scoring drop-outs were studied, in order to ascertain why the predictions based on test scores were not even better than they were, in these cases financial need, family pressures, and other non-aptitudinal factors seemed to be sufficient cause

This study has been criticized by Mursell (560 233) because the predictive value of the Seashore tests has generally been assumed to have been demonstrated by it, whereas the often referred to evidence is actually not based solely on the Seashore tests As Mursell pointed out, the data

were not presented in a way which made possible a definite evaluation of the predictive value of the Seashore tests, although this could easily have been done The value of the "cumulative key" may have been due largely to the intelligence test or to the ratings of previous music teachers There is implicit in Stanton's report, however, evidence to the effect that such data were available (748 68), and while it is true that if they were available they should have been reported, statements to the effect that the "lowest musical talent students were very short-lived in the school" should be taken into account No correlational data have been located, but in an earlier study (717) it was reported that in the four years, 1923–26, the percentage of students making grades of A, B and C on the music tests rose from 79 to 92, and teachers' estimates of student talent rose from 67 to 88 percent in the same categories The indication is that the higher level of talent revealed by the tests was confirmed by teacher evaluation While the reports are to be criticized for their lack of details from which generalization would be possible, it seems that the findings are not to be dismissed as completely as Mursell suggested they should be

Occupational differences were also studied by Stanton in a comparison of the scores of professional and amateur musicians with those of beginning students of music and non-musicians The former were found to be significantly higher than the latter, result which, in view of other findings already mentioned which showed that the test scores are not affected by training or experience, demonstrates the ability of the tests to differentiate the more talented from the less talented musicians

Preferences for different types of music were ascertained by Fay and Middleton (250), working with 54 college students Twelve musical selections were played to this group, and were rated by them for preferences They found that those who preferred classical music made higher scores on the pitch and rhythm tests than did those who preferred light classical music or swing, and also scored higher on the time test than did the swing fans If confirmed by other studies with larger groups and more extensive sampling of musical tastes this would be an indication of the role of musical aptitudes, for apparently the most "high-brow" music does appeal more to those who are best endowed It would be interesting to know what the relationship is between score on the Seashore tests and satisfaction with employment as member of dance and symphony orchestras, assuming that extraneous factors such as working hours, rates of pay, and employment stability could be controlled.

Use of the Seashore Measures of Musical Talents in Counseling and Selection From the preceding discussion it is apparent that the Seashore tests measure aptitudes which are relatively independent of mental ability and of each other, and that these are physical capacities which mature by about age 15 and are not affected by training or experience Although it is possible that one or two of them are in reality a combination of some of the others, the conclusion concerning their physical basis still holds Seashore's recommendation that the scores be used separately, and never combined, should be followed if musical capacities are to be meaningfully studied

The occupational significance of the Seashore tests is primarily musical, although they have been found to have some value in selecting persons for other jobs in which ability to make auditory discriminations seemed important It is doubtful whether they will ever have guidance values, however, outside of the field of music In it, it has been demonstrated that those who make high scores are more likely to complete training and to achieve professional status than are those who make low scores

In schools and colleges these tests can be used to advantage to screen out students who have musical talents which are often unsuspected or undetected, thus making it possible for them to develop their abilities for their own enjoyment and that of others, if not actually as a means of earning a living If the training and experience in music is found to hold a challenge, and if the skill acquired by the student seems equal to his promise, then it may be appropriate to consider vocational possibilities in music In schools of music the tests can well be used as a selection device, with due recognition of the fact that what a student has done with his musical ability by that time is at least as important a predictor of success as the ability itself Talents may be a *sine qua non*, but they cannot be sufficient in and of themselves

In guidance and employment centers the tests probably have value only in cases in which the prospect of further training is to be considered Job seekers who are already trained can best be judged on the basis of performance, that is, by means of auditions Those with some training but seeking more should also have auditions, in which the amount of previous training is taken into account by experienced teachers of music, but in such cases the talent tests should be of value in checking up on the trainability of the candidate It should probably be kept in mind, in such instances, that there are hierarchies in music as in other fields, and that some persons of lesser aptitudes may find ways in which to use

them whereas others with more aptitude may find doors closed For example, the potential night-club crooner may succeed with a modicum of talents assisted by good looks and a smooth manner, whereas a more gifted person who aspires to symphonic work may find himself outclassed in that field

Business and industry have so far apparently failed to find, or to attempt to find, any uses for these tests Perhaps certain types of machine tenders, inspectors, and mechanics, who need to judge the operation or defects of machinery by pitch or other auditory senses, could be selected partly by these means The hypothesis would first need to be validated, and then experimentation might actually find that thresholds are low enough so that selection on this basis is unnecessary When accident rates are relatively high in such jobs, however, it might well be worth experimenting with some of these tests A good automobile driver, for example, drives partly by ear, and responds at once to any change in the pitch of the customary noises of his machine, thereby forestalling some types of mechanical failure

CHAPTER XIV

CUSTOM-BUILT BATTERIES FOR SPECIFIC OCCUPATIONS

THE realization of the fact that tests are likely to give better predictions when designed and validated for a specific rather than for a general purpose has, for many years, led psychologists concerned with the selection of persons for professional training to devise batteries of tests for specific occupations Some of these have been designated as tests rather than batteries, and they have in general been called tests of professional aptitudes, hence names such as the Moss Medical Aptitude Test and the Ferson-Stoddard Law Aptitude Examination But they have actually been batteries of tests even when combined in one booklet, and they have generally, but not always, been designed for use in selecting professional students rather than in counseling students or selecting employees

This latter point is an important one, for many school counselors lacking a sound foundation in psychological measurement expect, on hearing of the existence of an instrument such as the Medical Aptitude Test, that they will find it invaluable in counseling their students or clients In general, those who press the matter are disappointed, for they often find that the desired test is used exclusively by the professional schools which developed it as a selection device, or that it is disappointingly like certain other familiar tests and therefore difficult to accept as a test of "medical," "nursing," or "teaching" aptitude

Whether available for general use, like the Engineering and Physical Science Aptitude Test, or restricted to use in professional schools, like the Medical Aptitude Test, batteries of tests for specific occupations are nothing more than combinations of existing types of tests of special aptitudes, usually modified in order to give them some of the specific predictive and face validity which is characteristic of the miniature-situation test Thus the Engineering and Physical Science Aptitude Test is made up of parts of the Revised Iowa Physics Aptitude Test, the Moore

Test of Arithmetic Reasoning, the Bennett Test of Mechanical Comprehension, and the Moore-Nell Examination for Admission to Pennsylvania State College, no special attempt was made to give the test face validity, presumably because mathematical and mechanical items have enough inherent face validity for technical fields The Coxe-Orleans Prognosis Test of Teaching Aptitude, on the other hand, is made up of especially developed items, such as vocabulary, information, and judgment But these items were selected or devised so as to have special bearing on education the vocabulary deals with subjects with which people who are interested in teaching are presumed to be familiar, sometimes verging on "pedaguese", the information is of a type which a would-be teacher might well be expected to possess, and the judgment items deal with classroom situations, behavior problems, and other matters in the handling of which a prospective teacher should presumably have some ability They certainly possess face validity, although whether they reproduce the life-situation on a small scale is of necessity an open question until experimentally demonstrated

One other type of battery of tests for specific occupations has recently been developed, one by the United States Employment Service's Division of Occupational Analysis and the other by the Psychological Corporation, these are respectively known as the General Aptitude Test Battery and the Differential Aptitude Tests, discussed in some detail in the next chapter The principle underlying this type of test battery is that, since each mensurable aptitude is usable in a number of occupations, standard instead of custom-built test batteries can be constructed and normed in such a way as to yield scores for a number of specific occupations This is fundamentally the same concept as that underlying the Primary Mental Abilities Tests but the approach is different Instead of beginning with a series of tests designed to measure the currently known and isolable aptitudinal factors and proceeding to ascertain their vocational significance, as in Thurstone's work, the procedure has been to develop tests which are fundamentally the same as those which have been demonstrated to have occupational significance, and then to obtain occupational norms for this uniformly developed and standardized series of tests Since mechanical comprehension tests have proved valid for some occupations but not for others, such a test is likely to be included in such a battery and given a weight in the score for a given occupation which is proportionate to its correlation with success in that occupation Sometimes, as in the case of the USES battery, the tests are parts of well-known tests

or close approximations of them, in other batteries, as in that of the Psychological Corporation, they utilize somewhat more original types of items designed to measure the same factors or constellations of factors as existing tests, in neither case is any attempt made to measure pure factors, as in Thurstone's batteries The USES does not use all of the tests in its battery for each occupation, selecting, instead, the few which have the most predictive value for any one occupation, the Psychological Corporation, on the other hand, has planned its work around the battery as a whole

The multi-occupational approach of the last two test batteries represents a new trend, different from that of the professional aptitude tests discussed in this chapter It results in one relatively brief series of tests with many applications, rather than in a collection of diverse test batteries, each usable only for one occupational field It is potentially much more valuable to vocational and educational counselors than is the professional aptitude test, for, with one battery of tests, it becomese possible to explore a great variety of occupational possibilities It takes time to accumulate occupational norms for such a battery of tests (the General Aptitude Test battery came into tentative practical use by the USES only in 1947, after nearly a decade of work, and the Differential Aptitude Test Battery, with the expenditure of $75,000, is just beginning to develop occupational norms), it takes even more time to develop special batteries for a number of occupations But it is also true that special occupational batteries are likely to have greater immediate validity for selecting students or employees than general aptitude test batteries, because of their miniature-situation elements and their custom-built character, these advantages are soon lost by the changes which take place in specific details, outmoding many miniature type items, and by the variations from one employing agency to another unless continuous research maintains the tests For example, the writer developed a personality inventory for the selection of Air Force pilots during World War II (801), which had more validity than the standard personality inventories and tests which were tried out at the same time, it was truly custom-built, with items phrased in the language of aviation cadets and content drawn from their wartime experiences, both actual and anticipated But changes connected with the end of the war made this test currently useless as a personnel instrument The obvious conclusion is that tests with *custom-built items* are best for *selection programs* in which conditions are relatively stable and investments are great enough to warrant the continuous

validation of existing tests and the constant construction of new instruments, but that for *counseling* purposes tests consisting of *generalized items* with occupational norms are the only practical choice

The tests discussed in this chapter are largely custom-built and were designed for personnel selection Those in the next are batteries of tests containing generalized items lending themselves to custom-built norming, designed primarily for counseling or for selection in programs unable to support a continuous program of test construction

As indicated above, tests of so called professional aptitude have almost invariably been developed for the selection of students in professional schools Professional training institutions invest so much in their students as to make selection essential, in a few instances they have been developed for the selection of other types of trainees or employees, but here also the investment in the trainee or worker has generally been large, as in the Air Force pilot-training program The tests have generally been kept confidential in order to prevent coaching, being made available only to member schools or official testing centers Tests of this type are briefly described in this section, as the great majority of users of psychological tests need no more than a knowledge of their existence and nature A few tests of this type are available for general use, and while these are discussed at slightly greater length they are not treated in detail because most of them have not been widely studied Both types are taken up under the title of the occupation for which they were developed, the occupational titles being arranged in alphabetical order

Business Executives Although little has been published on the subject in psychological journals, a great deal of time and money is currently being spent on the application of psychological methods to the selection of executive personnel General discussions of the executive selection and evaluation services offered by consulting organizations have been published in the May-June, 1946, issue of the *Journal of Consulting Psychology*, but evaluative studies are lacking on this very important phase of personnel psychology In general, there may be said to be five current types of work in executive selection and evaluation 1) the development of custom-built batteries of tests such as the Cleeton-Mason Vocational Aptitude Examination and the U S Civil Service Commission's experimental battery, discussed below, 2) the validation of standard tests for this particular purpose, as in the University of Minnesota's College of Business Administration project also discussed below, 3) the development of single tests for executive interests or other traits, best illustrated by

Strong's work with executives and public administrators, mentioned below, and discussed in connection with that inventory, 4) the clinical use of interviews and tests as commonly done by consulting psychologists, considered in this section, and 5) the use of clinically evaluated situation tests as developed by the British War Officer Selection Boards and carried further by the U S Office of Strategic Services for the selection of personnel for critically important assignments, also considered in this section despite the fact that it has so far not been written up as a procedure for the selection or evaluation of executives in business and industry

The Cleeton-Mason Vocational Aptitude Examination (McKnight and McKnight, 1947), is designed to measure aptitude for four types of business activity, clerical, accounting, administrative, and technical It is one of the few tests which purport to measure aptitude for executive work, it consists of eight subtests, the contents of which measure general information, arithmetic reasoning, analogic reasoning, reading comprehension, interest (as in Strong's), personality (as in Bernreuter's), vocabulary, and ability to estimate such things as the number of cars in the United States Although the authors have written a monograph on executive ability, in which they have analyzed the nature of the executive's task in a helpful manner, data on the validity of the *test* are so lacking as to make the test itself of little value in vocational counseling The purposes it might serve are probably better served at present by batteries of tests, such as the Otis Tests of Mental Ability, the Minnesota Clerical Test, and other tests of special aptitudes which have been rather thoroughly studied, except perhaps when the test items are completely tailor-made

A battery for the selection of public administrators has been developed by Bransford, Mandell, and Adkins of the U S Civil Service Commission (117,505), utilizing two standard tests of intelligence (the A C E Psychological Examination and Thurstone's Estimating Test) and custom-built tests of current events, data interpretation, administrative judgment, and knowledge of agency organization and personnel The criterion of success was a combined rating of administrative effectiveness, the average number of raters per employee being four The top management ($6,200 to $10,000) group consisted of 20 persons, for this group, the correlations between criterion and A C E were 64, Current Events 64, Interpretation of Data 65, and Administrative Judgment 68, other validities for this group were low For the staff group (63 specialists at $2300 to $7500) the validities were 30, 26, 41, and 49 The multiple

validity coefficient for the staff group (the only one large enough for computation) was .55 These data suggest that truly custom-built tests of executive ability may have considerable validity, but this battery cannot yet be considered to have been validated, in view of the fact that the persons tested were already on the job at the time of testing Its validity can be considered established only after applicants for employment have been tested and followed up This is especially true of custom-built batteries, some of the items of which may be more readily handled after one has worked in the situation than before But, as the test authors concluded, this preliminary work with the battery suggests that it may have merit and that further validation should be carried out

The validation of a battery of standard tests for the selection of students of business administration at the University of Minnesota was written up by Douglass and Maaske (207) This battery was designed solely for local selection purposes, but the investigation does provide some suggestions as to what types of tests are likely to have predictive value The tests which showed the closest relationship to success in the college of business administration measured knowledge of social terms (Wesley College Test of Social Terms) and of business mathematics, with correlations with first-year honor point ratios of .56 and .47, respectively, and an R of .64 It need hardly be pointed out that success in training may be much more dependent upon academic ability (the verbal factor) than success on the job, and that the selection or upgrading of executives might require a rather different battery of tests

Strong's attempts to develop scales of executive interests (770,779) have shown that executives are not a homogeneous occupational group, but actually an extremely heterogeneous one, drawn from a great variety of fields such as sales, accounting, engineering, clerical, and skilled occupations Under these circumstances it seems probable that the traits which executives have in common are fewer and more difficult to isolate than those which subdivide the group It might, for example, be easier to distinguish insurance executives from insurance salesmen, engineering executives from engineering technicians, or office managers from office clerks, than to distinguish executives as a group from a group of men-in-general which includes insurance salesmen, engineering technicians, and clerks Strong's work has shown that, in the field of interests at least, what the salesmen, technicians, and clerks have in common is what the insurance executives and office managers have in common The lines are drawn vertically rather than horizontally, the executive salesmen

being the most able salesmen, the executive engineers being the most able engineers, the executive office workers being the most able office clerks In the field of aptitude, also, being an executive may be a matter of being superior in one's field, rather than having notable characteristics which are common to all types of executives (abstract intelligence would be an exception to this statement, in that executives in a given field and in all fields could be expected to excel in such a general ability)

A battery of standard tests administered to 15 superior and 10 average executives of a firm of consulting management engineers by Thompson (826) is of interest as one of the few published studies reporting positive results The tests used included the Wonderlic Personnel Test, Michigan Vocabulary Profile Test, Cardall Test of Practical Judgment, Kuder Preference Record, Adams-Lepley Personal Audit, Beckman Revision of the Allport A S Reaction Study, Guilford-Martin Personnel Inventory, and Root I-E Test The criterion consisted of performance records (not described) and ratings by partners, how reliable these were is not stated Differences between the superior and average groups, significant at or above the 5 percent level, were found with the Wonderlic, Michigan Vocabulary (Government, Physical Science, Mathematics, and Sports subtests), Kuder (Mechanical and Social Service), and Adams-Lepley (Firmness and Stability) tests Both groups were found also to be above the 93rd percentile on the Kuder Persuasive scale All of the reported differences favored the superior executives, except that on the Kuder Social Service scale in this characteristic the average or less successful executives were at the 79th, while the more successful executives were at the 51st, percentile These results portray the successful management engineer executive as superior to less successful partners in mental ability, technical and governmental vocabulary, sports vocabulary, mechanical interests, firmness, and stability, and inferior in interest in social service As Thompson's groups were very small these conclusions are highly tentative, cross-validation might change the picture considerably Further studies of this type appear, however, to be worth making

The clinical use of interviews and tests is perhaps the most common method now used by consulting psychologists in the selection or evaluation of executive (and sales) personnel Although it does not make use of a total score based on a test battery, this procedure is briefly described here because of its prevalence and because it constitutes one method of using tests

Flory and Janney (267) have listed five factors which experience has

led them to believe must and can be appraised in executive evaluation: intelligence, both abstract and concrete, emotional control, defined as ability to maintain steady output without emotional tension under varying and trying circumstances, skill in human relations, or leadership in face-to-face situations, insight into human behavior, both one's own and that of other persons, and ability to organize and direct the activities of others Sóme of these traits can be rather effectively measured, intelligence, for example, by means of standard tests and perhaps by certain Rorschach indices But others, such as emotional control and insight into human behavior, have not as yet lent themselves to effective measurement The judgment of such qualities is a much more complex and unreliable procedure than the statement by Flory and Janney implies

The procedures used by the consultants in question consist of a detailed personal history secured in an interview lasting from twenty minutes to two hours, "suitable objective instruments" to probe areas of adjustment, and a clinical interview for the checking of symptoms revealed by the personal history and the tests Fear, in part of another article in the same symposium (2) mentioned comparable methods used by another organization, without going into details other than stating that they can be used only by a highly trained psychologist

This procedure is nothing more than that used by any well-trained and balanced user of tests for selection purposes it consists of selecting and interpreting the results of tests believed likely to throw light on significant aspects of the applicant's qualifications, gathering important supplementary data by other means, and synthesizing them into a meaningful picture But, in contrast with test procedures for many other types of work, it is actually *less* than what is done in most personnel evaluation programs For in the best use of tests in personnel selection and evaluation the tests have been previously subjected to experimental validation for the work in question, and are used because there is an objectively demonstrated relationship between the test score and success in that job, whereas in the procedure under discussion few if any such relationships have been established and the additional clinical work is an attempt to make up by subjective procedures for what has not been done by objective techniques Flory and Janney's "suitable objective instruments" for probing personality may be objective in form, and suitable in the best judgment of a competent vocational psychologist, but the existence of a relationship between scores on such tests and success in executive work

had not, at the time of their writing, been demonstrated The personnel selection and evaluation procedure described by Flory and Janney, Fear, and others is a clinical procedure which uses tests *diagnostically* but not *prognostically,* the predictions are based on *clinical judgments* and not on the *known relationships* of *tests*

To underline this fact is not to deny the value of current psychological methods of executive selection as a matter of fact, they are probably superior to other presently available methods It is merely to point out a major difference between the use made of tests in such programs and in most other selection or evaluation procedures The reasons for this difference are clear they lie in the elusiveness of personality factors, in the primitive state of development which characterizes our present methods of appraising personality characteristics, and in the fact that executive selection is so vitally important that it justifies the time of the vocational and clinical psychologists who must make the clinical judgments involved Subjective and even defective though these judgments may be, they represent the best available informed guesses are preferable to uninformed guesses, and better-informed to less-well-informed In the equally complex problem of predicting success in pilot training, for example, it was found that judgments made in psychiatric interviews with aviation cadets had a correlation of only .27 with success in flying training, as contrasted with a validity of .66 for a custom-built and objectively scored test battery, the psychiatric interviews were of little more than chance value, and much less effective than the test battery, but if no such battery of valid tests had been available the weeding out of even a few failures would have justified depending upon the clinical judgment of the psychiatrists The suggestion emerging from this discussion is that it would be well worth the while of organizations interested in the selection and upgrading of executives to finance whatever fundamental research is a prerequisite to the development of better tests for the measurement of characteristics which may affect success in administrative and top-managerial work

Clinically evaluated situation tests used by the Office of Strategic Services have been described by Murray and MacKinnon (558) and by the Assessment Staff (33) In this work they were concerned with appraising "the relative usefulness of men and women who fell, for the most part, in the middle and upper ranges of the distribution curve of general effectiveness or of one or another special ability," and with assessing a number of "personality qualifications—social relations, leadership, dis-

cretion " As it seemed that none of the conventional screening devices tested good will, tact, teamwork, freedom from annoying traits, leadership, and other social qualifications, special procedures had to be devised In other words, the project had to develop methods of appraising executive ability, since none were available, that the executive ability was to be applied in "cloak-and-dagger" work is incidental, and should not blind civilian personnel workers to the possibilities of the methods tried That they are also being used in the British civil service is further testimony to their general promise

The O S S procedure consisted essentially of bringing about 18 candidates to a house party for a period of three and one-half days The activities of the house party were directed by a staff of psychologists, psychiatrists, and sociologists Data were gathered by means of casual observations, standard tests of intelligence, mechanical comprehension, etc , projective tests such as Incomplete Sentences, Thematic Apperception, and the Rorschach used primarily to assess motivation and emotional stability, personal history interviews of an hour and one-half, group situation tests, one requiring working with a team to accomplish a feat of physical prowess, another a discussion, in both of which leadership might develop, and some assigned leadership problems in which the examince must lead his group, individual situational tests involving frustration-tolerance and a stress-interview, an obstacle course, tests of observing and reporting details, tests of propaganda skills as shown in the preparation of a pamphlet to disturb Japanese workers in Manchuria, psychodrama involving difficult social situations, debate in a convivial party, a sociometric questionnaire concerning fellow candidates, and judgment of others as revealed in sketches of the five men known best during the three and one-half days

Data obtained by these methods were clinically evaluated by the staff subgroup responsible for the study of several candidates, and reworked in case conference by the whole staff About 20 percent of the 5,500 men and women thus studied were not recommended for duty, 1,200 of those who went overseas were followed up and evaluated by supervisors and three or four associates The choice and collection of criterion data was not undertaken until late in the war, and convincing quantitative validation proved especially difficult (33 Ch 9) Despite these difficulties a validity coefficient of 39 was obtained for a sample of 31 candidates assigned to appropriate duties The authors conclude, with some justification, that the true validity of their procedure was probably between 45

and 60 (33 424) Two important points can now be made on the basis of this work

The first is that the possibility of following up employees and obtaining evaluations under even the most difficult circumstances is rather conclusively demonstrated by the obtaining of evaluations of men and women who were appraised in this country and followed up in scattered combat areas,

The second is that there are many devices for obtaining potentially significant and quantifiable personality data which psychologists have only begun to explore, making the field of personality measurement a rich one in which to carry on research Since executives play crucial roles in their organizations, and represent considerable investment of company or public funds, the exploration of these possibilities should be well worth the while of business, industry, and government

Dentists Research in the selection of students for dental schools was summarized in 1940 by Bellows (65) Most of the batteries used consisted of standard tests selected because it was thought they would have validity for this purpose, but two included tests which were developed specifically for dental selection One was the *Iowa Dental Qualifying Examination* of the State University of Iowa (724), the other a battery developed at the University of Minnesota partly on the basis of the Iowa work (208) The Iowa tests were information on the development of the teeth, reading comprehension (dental anatomy), memory for nomenclature, predental chemical information, predental zoological information, a worksample (trimming a plaster of Paris block to specification), and a paper-and-pencil test of spatial relations The correlations between scores on the first five tests and theory grades in thirteen dental schools ranged from 11 to 74, the average being 53, for the worksample the correlation with grades in first-year technique courses was 62, that for the spatial test was 41 Several possible combinations of tests were used in the Minnesota studies (208), their validity varying somewhat not only from battery to battery but also from year to year, the numbers varying from 83 to 111 One battery consisted of predental grades ($r = 45$), a metal-filing worksample (53), the Iowa Visual Memory Test for Nomenclature (40), the O'Connor Finger Dexterity Test ($-$ 40), and the Iowa Spatial Relations Test (52), the multiple correlation with total grades in dental school was 78, even when only the filing, memory, and dexterity tests were used When laboratory (Prosthesis) grades were used as a criterion, the Metal Filing Test (custom-built) had a validity of 60 while that of

the Finger and Tweezer Dexterity Tests (standard) was − 35 and − 43 (high time scores are bad, hence the negative relationship)

These studies show that grades in dental school have been predicted with considerable success by means of batteries of tests, some of which were constructed especially for that objective However, the value of this approach can be judged only by comparing its results with those of studies which have used standard tests weighted for dental selection on the basis of local validities Only Harris' study (341) permits such a comparison, made in a different school with a criterion (grades) which may have been more or less reliable than those in the Iowa and Minnesota studies his multiple validity coefficient, using predental grades and intelligence test as predictors, was 67, which is substantially lower than the 79 obtained at Minnesota with a special battery Whether or not the additional validity justifies the extra labor of constructing the special battery depends, of course, upon the expense of the mistakes which result from using an inferior selection procedure

Engineers Although various investigators and institutions have developed procedures for the selection of engineering students, and the Engineers Council for Professional Development is now working on a large-scale study of this type (28 8), no so-called tests of engineering aptitudes were published until the appearance on the market of the *Engineering and Physical Science Aptitude Test* (Psychological Corporation, 1943) Oddly enough, this is not, at least at present, a test for selecting students for colleges of engineering It was developed in connection with the war-industry training program at the Pennsylvania State College, and so has norms for miscellaneous young men and women, some of them not high school graduates, who applied for technical training at the trade and technician level in connection with war industries This test, or rather battery of tests, is not a test with custom-built items in the sense in which that term is used here Instead, it consists of items from existing tests of special aptitudes, selected on the basis of item validities to constitute a new battery The items were therefore custom-selected, but not custom-built, they are of possible general significance, rather than drawn from and restricted to the local situation It is only the weights and norms which are custom built

The tests from which the items were selected on the basis of local validities were the Iowa Physics Aptitude Test (revised), which provided the Mathematics, Formulation, and Physical Science Comprehension Tests, the Moore Test of Arithmetic Reasoning, which supplied the

Arithmetic Reasoning Test, the Bennett Test of Mechanical Comprehension, from which came the Mechanical Comprehension Test, and the Moore-Nell Examination for Admission to Pennsylvania State College, vocabulary section, which provided the Verbal Comprehension Test These are all tests which have been found to have some value in predicting success in technical and engineering courses, but until engineering norms have been provided for this form of the test it must be considered more likely to be dangerous than helpful in selection or counseling A high school senior might, for example, compare very favorably to the norm group of miscellaneous young men and women, some of whom did not have the academic ability to finish high school, but find it difficult to compete with typical college freshmen (confirmed by a study by Fagin at Brooklyn Polytechnic Institute) On the other hand, since the items have all been twice selected on the basis of validity for predicting success in technical training at some level (in the original test and in this battery) the battery should be a very promising one on which to collect local data and to establish local norms An engineering or technical school which cannot at once invest much money in test construction and validation would probably find that this battery provided a ready basis for establishing local selection criteria

As currently available information concerning the Engineering and Physical Science Aptitude Test is limited to the original study and is contained in the manual and in the article by Griffin and Borow (314), work with it is not discussed here in any detail It should suffice to say that correlations between scores on this test and grades in technical courses ranged from 13 to 71, depending upon the course and the subtest, and that the correlation between total score and average grade was 73 Subtests showed higher correlations with grades in the types of courses with which one would expect them to be related than in others the correlation of 71, for example, was between the Mathematics score and grades in mathematics, whereas a correlation of 14 was found for Mathematics score and grades in a course in manufacturing processes

Attempts to develop batteries of tests for selecting engineering students, in which standard tests have been used as *tests* rather than as *sources of items,* are perhaps best illustrated by studies conducted by Holcomb and Laslett (375), Laycock and Hutcheon (456), and Brush (122) Holcomb and Laslett used the MacQuarrie Mechanical, Stenquist Picture, and Stenquist Assembly Tests, and the Strong Vocational Interest Blank (engineer scale) They computed no multiple correlation coefficients,

although they found validities of 48, 15, 43, 16, and 32 respectively

Laycock and Hutcheon used the National Institute for Industrial Psychology (England) Form Relations Test, the Cox Mechanical Aptitude Tests (Models and Diagrams), and the physical science score on the Thurstone Interest Inventory, together with high school grades and scores on the A C E Psychological Examination The best combination of this group consisted of marks in grade 12, A C E score, Form Relations, and Physical Science Interest, the multiple r being 66

Brush's study included the Minnesota tests, the Wiggly Block, the Cox Mechanical Aptitude Tests (Explanation, Completion, Models), the MacQuarrie, the Thorndike Intelligence Examination, and the Columbia Research Bureau science tests, not all students took all tests, as he worked with two groups The multiple correlations for all tests and four-year engineering grades were 54 for one group (no intelligence test included) and 61 for the other (including intelligence test) With the first group the best battery was probably that consisting of the Minnesota Paper Form Board and the Cox Models, with an R of 46 For the second group the best batteries were one consisting of the Thorndike, C R B Algebra and Geometry, Cox Models and Completion, Minnesota Paper Form Board and Interest Analysis, with an R of 59, another consisting of the C R B Physics, Chemistry, Geometry, and Algebra Tests, for which the R was 585, and a third made up of the Thorndike, C R B Algebra, Cox Models, and Minnesota Interest Analysis, with an R of 59 The highest correlations for single tests were for C R B Algebra and Physics, and Thorndike Intelligence, respectively 51, 50 and 43

It will be interesting and valuable to see, at some future date, the relative validity of a battery such as the EPSA Test, in which items have been custom-selected, when compared with batteries of standard tests such as these

Lawyers Tests and test batteries for the selection of law students have been developed at a number of universities, notably California, Columbia, Iowa, Michigan, Minnesota, and Yale, and most recently by the Educational Testing Service (28 8), a recent review of work with these and other tests in law schools has been prepared by Adams (4) The pioneer test in this field appears to have been the *Ferson-Stoddard Law Aptitude Examination* (West Publishing Co, St Paul, Minnesota, 1927) It consists of four parts a reading comprehension and recall (after the other parts) test based on a law case, a reading comprehension and reasoning test based on another case, a verbal reasoning test, and a reading

comprehension test based on legal material The test has been used only by law schools and has not been available to counselors Ferson and Stoddard (254) found that the Law Aptitude Examination had a correlation of 54 with the first-year law grades of 100 students at the University of Iowa, as summarized by Adams (4), subsequent studies of the test yielded validity coefficients of 54 at Tennessee, 34 at Newark, 42 at the New Jersey Law School, 46 at Illinois (first semester only), and 49 at Chicago All of these studies agree, then, in showing considerable validity in the test, about that shown by scholastic aptitude tests in liberal arts colleges, since the populations of law schools are somewhat more homogeneous than those of colleges, the test presumably has somewhat more validity It is therefore interesting to compare its validity, in the last study, with that of the A C E Psychological Examination, which was found to be 56 in contrast with that of 49 for the Ferson-Stoddard Law Aptitude Examination As the combined tests yielded a correlation of 62 it seems that, although they were not measuring exactly the same thing, the contribution of the professional aptitude test was not great In the Illinois study (916) Welker and Harrell found that pre-law grades had much more predictive value than the aptitude test (75 as compared to 46), and that the correlation for the combined indices was not much higher (78) Only three of the six scores of the Ferson-Stoddard test (Part 2 of which yields three scores) were found to have any appreciable correlation with grades these were Part 2C, Relevant Facts, Part 3, Logical Inferences, and Part 4, Matching, these validities were 17, 28, and 31, respectively, the validities of A C E part scores were of the same order, but more consistently so The implication is that a good general intelligence test is at least as useful as this professional aptitude test, especially when one notes, with Welker and Harrell, that the effective law aptitude subtests are the reasoning rather than the "legal memory" tests Studies at the University of Minnesota (206) obtained correlations of custom-built tests with law grades which were as good as those for intelligence tests, but multiple correlations permitting comparison were not reported

In 1943 Adams (4) published studies of a new *Iowa Legal Aptitude Test*, developed for use in the same institution as the Ferson-Stoddard nearly twenty years earlier Its preliminary form consisted of eight subtests, the first three of which are not legal in content, while the last five are Part 1 is a verbal analogies test, Part 2 is a mixed relations or more complex analogies test, Part 3 contains opposite items of a verbal type,

Part 4 is a test of memory for material in a judicial opinion read before Part 1 or two hours earlier, Part 5 is a reading comprehension test stressing judgments of relevance, Part 6 is also a reading comprehension test adapted from Part 2 of the Ferson-Stoddard Test, Part 7 is a verbal reasoning test, and Part 8 is a legal information test When the first-semester grades of 110 law students were correlated with part and total scores on the Legal Aptitude Test, the former were found to range from 36 (Part 5, reading comprehension for relevance) to 57 (Parts 3 and 7, verbal opposites and verbal reasoning), while the validity of the total score was 65 It was decided to use Parts 3, 7, and 8 (verbal opposites, verbal reasoning, and legal information) in the final form of the test, the multiple correlation of these subtests with the criterion was 67, higher than that of the total score on the preliminary form of the test Although no comparisons were made with general intelligence tests, comparison with the predictive value of achievement tests and pre-law grades indicated that in this case the professional aptitude test had more validity than the non-specialized indices This was presumably because the professional aptitude test was, itself, a highly refined test of general intelligence, couched in terms most appropriate to the field in question, to which was added an interest-achievement factor by the inclusion of a subtest of legal information

Nurses Batteries for the selection of nursing students have been developed by a number of university schools of nursing, and by independent organizations or individuals working on a consulting basis with nursing schools

The George Washington University Series of Nursing Tests (Center for Psychological Service, George Washington University, 1944) was developed from the Moss-Hunt Nursing Aptitude Test, first published in 1931 and available to counselors The series incorporates a modified form of the Nursing Aptitude Test, consisting of five parts, as follows judgment in nursing situations, memory for anatomical diagram and nomenclature studied during the test, nursing information, scientific vocabulary, and following directions in filling out a nurse's report form This test is, obviously, custom-built as to items, drawing heavily on the technical content of nursing as it might have been experienced before training or as presented in the test itself A second test in the series is a Reading Comprehension Test, utilizing material from commonly used textbooks in nursing schools The third test is an Arithmetic Test, the fourth is a General Science Test based on high school courses, and fifth is an

Interest-Preference Test somewhat resembling Strong's Vocational Interest Blank, the items of which were selected because they differentiated nurses from non-nurses Norms based on high school graduates applying for admission to nursing schools are provided with the manual, together with suggested critical scores for interpretation, but it is recommended that local norms be developed because of differences in standards No indication of the numbers on which the standardization was done is included, nor are there any data on the validity or reliability of the new series of tests No references to them have been located in the literature Although the test items look promising and have obviously been based on the best available experience in nurse selection programs, validation data such as have been provided by other investigators using the earlier form of the Aptitude Test for Nursing are needed In these studies of the earlier form of the first subtest of the present series Douglass and Merrill (209) found correlations ranging from .54 to .62 with grades in the first year of nursing school at the University of Minnesota, and Williamson and others (929) found correlations of .34 and .37 with grades in twenty schools of nursing As these grades were very unreliable in some schools the validity seems lower than it actually was, in one school with a better marking system the validity was .49 It seems clear that this one part of the present series is what the manual suggests a "specialized intelligence test for prospective nurses" The other subtests appear to be specialized achievement and interest measures for prospective nurses, but need to be evaluated as such

The Nursing Entrance Examination Program of the Psychological Corporation has developed another battery of tests for use in schools of nursing This battery is administered periodically at various centers throughout the country, by arrangement with co-operating institutions, it is not available for general use Unlike the battery developed by Hunt, it consists of standard tests found useful in selecting nursing students rather than of custom-built tests, it is only the norms that are custom-developed The program has been described by Potts (611)

Other standard tests have been used in studies referred to earlier (209, 929), conducted at the University of Minnesota and co-operating schools of nursing In these it was found that standard tests of vocabulary (Co-operative Test Service), English, and General Science had substantial validities, as high as .44, .53, and .58 in one school where marking was reasonably reliable Douglass and Merrill found a validity of .77 for the Moss-Hunt Test of Nursing Aptitude and the Co-operative General

Science Test Crider (181) found that the Strong Interest and Bell Adjustment inventories added little to predictions based on the Otis Test of Mental Ability, confirming Douglass and Merrill's correlation of 20 for Strong's nurse scale and grades

Pharmacists Until recently little attention was paid to the scientific selection of students of pharmacy, and little was known concerning psychological factors related to success in this occupation During World War II, however, members of the occupation became more self-conscious as a profession, even to the point of changing the status of pharmacists in the Army from enlisted to commissioned grade Since the War the American Pharmaceutical Association has been engaged in a co-operative study with the American Council on Education, one of the purposes of which is to develop better methods of selecting pharmacists, and Schwebel (unpublished study) has developed a pharmacist scale for Strong's Vocational Interest Blank

Physicians *The Moss Medical Aptitude Test* (Association of American Medical Colleges, 1930) was for many years the standard instrument for the selection of medical students, used by most medical schools in the United States and not available to others New forms were provided periodically, but the content is rather like that of the Moss-Hunt Nursing Aptitude Test which has already been described and which was based in part on Moss's experience with the Medical Aptitude Test Parts deal with comprehension and retention, logical reasoning, scientific vocabulary, etc, making the test one of intelligence measured by means of medical material Some of the studies published in the Association's journal have shown that there is a tendency for high-scoring applicants to succeed in training and to be rated favorably as interns, whereas those who make low scores tend to do poorly Moss (550) reported that one percent of the top-decile students failed, as contrasted with 18 percent of the bottom-decile students Chesney (155) found that refusing to admit anyone in the lowest decile would eliminate 25 percent of the failing students, 15 percent of the mediocre students, 7 percent of the fair students, and only 3 percent of the good students But Douglass (205) and Cavett and others (152) found validities of only 12 to 34 for various classes at the University of Minnesota, compared to 40 to 57 for liberal arts grades Moon (535) found a closer relationship at Illinois, where the validity was 42 and liberal arts grades had a validity of 49 The Minnesota Medical Aptitude Test, another custom-built battery, had validities of only 14 to 40 Strong's Physician scale had a validity of only

16 for 131 students, using first-year honor points as criterion Stuit (784) obtained correlations of 23 and 32 between the Moss Test and first-year grades in medicine at the University of Iowa, as compared to correlations of 45 and 46 between college grades in liberal arts and science courses, on the one hand, and medical grades on the other The Moss Test and science grades yielded a multiple correlation of only 49 These studies suggest that, although the Moss and Minnesota Medical Aptitude Tests have some value in selecting medical students, they do not add much to predictions made on the basis of undergraduate college grades Apparently further study and development of new types of instruments is needed in this field In the meantime, the standard measures of intelligence and achievement in appropriate areas will probably prove as useful as the professional aptitude test in appraising promise in this field The Educational Testing Service (28 9) now handles this admission-testing program for the A A M C

Pilots Apart from embryonic efforts in the first World War, tests for the selection of aircraft pilots were first developed early in World War II by the Civilian Pilot Training Program of the Civil Aeronautics Administration, the work of which was summarized by Viteles (903), were further developed for the U S Navy under Jenkins' leadership (399), and especially by the Army Air Forces Aviation Psychology Program (214,265) under Flanagan The most far-reaching of these, both in the variety of tests used and in the extent of its validation procedures was the last named, as it included tests comparable to those originated by the other two programs, only it is described here

The Aviation Cadet Classification Battery (U S Air Force, 1942, revised in 1943 and subsequently) consisted of a personal history questionnaire arranged in multiple-choice form and stressing experiences and background factors which had been found related to success in flying training, two spatial orientation (perceptual) tests utilizing aerial photographs and maps, a reading comprehension test, a dial and table reading test involving taking readings from airplane instruments and aeronautical tables, two instrument comprehension tests also based on flight instruments, a mechanical principles test based on the Bennett, a general information test presumably tapping interests and personality traits underlying the possession of information found to be related to success or failure in flying training, two mathematics tests, a rotary pursuit (eye-hand co-ordination) test, a lathe-type two-hand co-ordination test, a stick-and-rudder test in which controls are moved to match light signals

appearing in a prearranged pattern, a rudder control test in which the examinee's seat is kept in equilibrium by movements of the rudder with the feet, a discrimination-reaction-time test requiring the selection of a switch to be moved in order to put out a series of lights, and a pegboard measure of finger dexterity (214) Most of these, it may be noted, involved custom-built items the biographical data items were written to tap aspects of experience which might be related to flying success, the perceptual items involved perception of the type used in pilotage, the eye-hand-foot co-ordination test used a stick and rudder, etc Although correlational analysis techniques were used to insure relative independence of the tests, the miniature-situation element was strong in most of them

As is necessary in a custom-built selection testing program in which conditions are constantly changing, these tests, their antecedents, and their successors, were continuously validated as data concerning new criterion groups were received The most impressive of these validation studies (214 Ch 5, 264) was made with a group of 1143 candidates for aviation cadet training who were sent to pilot training regardless of their scores on psychological tests Analyses were made to reveal the comparative validity of the psychological tests, the cadet selection battery as a whole, the Adaptability Rating for Military Aeronautics (psychiatric examination), the Army General Classification Test, the Aviation Cadet Qualifying Examination (custom-built intelligence test used in preliminary screening), and years of education Data are reproduced in Table 26 (p 350)

The correlations given are with success in training through advanced flying school, that is, with ability to win wings and a commission Outstanding in the above data are the following facts

The three most valuable tests are paper-and-pencil tests,

The most valid tests are custom-built even in item content,

The battery has more predictive value than the best single test,

Objective tests have more predictive value than psychiatric judgment

Later work with this battery has involved the factor analysis of these and certain other tests (316,317), the refinement of the most promising, the addition of subsequently developed tests to the battery, and, since the end of World War II, an ambitious joint project of the Air Force, Navy, and American Institute for Research in which a battery of paper-and-pencil tests is being developed which will measure with maximum economy all of the characteristics which have so far been found to con-

TABLE 26

RELATIVE PREDICTIVE VALUE OF CERTAIN CUSTOM-BUILT AND STANDARD PSYCHOLOGICAL TESTS AND CERTAIN OTHER INDICES FOR SUCCESS IN PILOT TRAINING (After DuBois)

	Test	*Validity*
Pilot Tests	General Information	51
	Instrument Comprehension II	48
	Mechanical Principles	43
	Complex Co-ordination (Stick-Rudder)	42
	Discrimination-Reaction-Time	42
	Spatial Orientation II	40
	Dial and Table Reading	40
	Rudder Control	40
	Two-Hand Co-ordination	36
	Biographical Data	33
	Stanine (Battery score)	66
	Aviation Cadet Qualifying	50
	Army General Classification	31
	Education	21
	Flying Adaptability Rating (Psychiatric)	27

tribute to flying success Studies were also made which ascertained the predictive value of the wartime battery and its components for success in combat (467), this was found to be significant, although attenuated by the relatively small and select group of pilots which reached combat and the complexity of the criterion the number of planes shot down by a fighter pilot in England in 1942 cannot be compared, for example, to the number shot down in the same theater in 1945 when air superiority had changed hands and daylight bomber raids were unknown

The American Institute for Research, established by Flanagan and other aviation psychologists on the basis of their wartime experience, has carried out a number of research projects for the Civil Aeronautics Administration and several of the commercial airlines, analyzing the work of the airline pilot and constructing a battery of tests for the evaluation of pilot proficiency which might be used in selecting personnel for commercial airlines The Institute has established testing centers at which the current form of this battery is now being used in such selection, but data concerning it have not yet appeared in the literature

Psychologists The post-war demand for clinical and vocational psychologists resulted in a great increase in the number of candidates for training in psychology and a strain on training facilities Graduate departments of psychology, the Veterans Administration, the U S Public Health Service, and the American Psychological Association worked together on the problem of improving the selection and training of psy-

chologists One result of this co-operation was a project carried out at the University of Michigan, under the direction of E L Kelly, for the development of a battery of tests for the selection of students for training in clinical psychology, another project provided for the study and revision of the psychologist scale for scoring Strong's Vocational Interest Blank under the direction of D G Paterson of the University of Minnesota

Salesmen More attention has been devoted by business and industry to the problem of selecting salesmen than to any other single group save possibly executives Unfortunately too many business concerns have been so near-sighted that they have been willing to employ psychological consultants for actual selection work but have not been willing to finance the research which should precede the development of any new method or instrument, whether it be psychological, chemical, or mechanical Even scientifically trained executives such as engineers often fail to realize that developmental work must be done in personnel selection just as in manufacturing And there have too often been psychologists and pseudo-psychologists available who were willing, either through ignorance of the complexities of personnel testing, or through eagerness to supplement academic incomes, to attempt to meet the needs of business and industry on their own inadequate terms So-called institutes for aptitude testing therefore flourish in most of our large cities, testing candidates for sales positions and making recommendations to referring employers which are based to an undetermined extent upon hunches and shrewd judgments made independently of the tests, and partly upon clinical evaluation of test scores, as described by Flemming and Flemming (266) and discussed in connection with executives, above

The Moss Test for Ability to Sell (Center for Psychological Service, George Washington University, 1929) is one of the few tests or batteries of tests marketed as a device for selecting salesmen It consists of items designed to test memory for names and faces, judgment in sales situations, observation of behavior, comprehension and retention of selling points in reading material, following directions in making out sales records, and sales arithmetic, and has norms based on department store salespersons Although it has been tried in numerous sales situations, the results have not generally been published in the journals The prevailing opinion of it among department store personnel workers known to the writer is not favorable

The majority of researchers who have experimented with test batteries

for the selection of salesmen have utilized personal history blanks, interest inventories, and personality inventories, as well as intelligence tests The first named are generally custom-built, the second is usually Strong's Vocational Interest Blank as a source of either a score obtained from a standard key or of items for the development of a new key, and the personality measures have included the Bernreuter Personality Inventory, the Humm-Wadsworth Temperament Test, or other well-known inventories For example, Bills (90) reported on the use of the life insurance and real estate salesman's keys of Strong's Blank, personal data, the Bernreuter, and a mental alertness test, the last two were of little value, but the others, combined, significantly improved the selection of successful salesmen Kurtz (449) worked with life insurance salesmen, using personal history items and Kornhauser's personality inventory and obtaining correlations of 40 with production Men who rated A had twice the chance of staying in the business for a year that men with E ratings had Similar findings have been reported with salesmen of more tangible things than life or casualty insurance Otis (580) used personal data items, a combination of Strong's life insurance and real estate keys and the Bernreuter, with salesmen of a detergent company, finding that the first two were effective predictors of success while the last-named test was not Building materials salesmen were studied by Ohmann (572), who used only personal data, he found a correlation of 67 between a questionnaire of 13 items and his most reliable criterion, annual commission earnings Viteles (902) tried the Humm Wadsworth Temperament Test with 59 appliance salesmen, but found that 12 of the 20 who had "desirable" patterns were discharged or resigned during the try-out period

From studies such as these, more thoroughly reviewed by Schultz (683) and by Kornhauser and Schultz (443), the conclusion to be drawn is that, contrary to the expectation of many personnel consultants, personality inventories have little or no value in the selection of salesmen The reasons for this will be discussed in a later chapter dealing with such instruments The most effective batteries have consisted of the sales keys of Strong s Vocational Interest Blank and, especially, custom-built personal history questionnaires The nature of the personal history items which prove valuable varies somewhat with the type of saleswork, but some consistent trends are revealed In Ohmann's study the 13 valid items were as follows height, age, marital status, number of dependents, amount of life insurance, debts, years of education, number of clubs and

organizations belonged to, years on the last job, experience in the line of sales in question, average number of years on all jobs, average monthly earnings on the last job, and reasons for leaving the last job It is notable that, although these salesmen were handling a tangible, building materials, the success of life insurance salesmen has also been found to be related to age, marital status, dependents, amount of life insurance, organizations belonged to, etc (91) Stokes (762), reviewing what experience has shown to be important in research in the selection of salesmen, has like others emphasized the need to take into account the job environment of the salesman, pointing up the fact that, despite the similarities which exist between sales jobs, and the more or less universal validity of Strong's sales keys, specific factors are found in any job which make custom-built batteries of tests more valid than standard tests His second point then follows of necessity research in the selection of salesmen must be dynamic, for it must continue to take into account the changes which take place in the environment in which the salesman is working and therefore in the demands of his job The fact that Strong's Vocational Interest Blank has been found to predict success in sales jobs, but in very few other occupations (see the discussion of Strong's Blank in Chapter 17), appears to confirm this point concerning the special importance of interest and motivational factors in selling

Scientists The importance of scientific occupations was emphasized as never before during World War II and its aftermath, when some countries such as Great Britain kept their science students and scientists draft-exempt because of their potential contributions to the war effort, and when the various Allies engaged in a scramble for the talents of the scientists of the conquered countries, particularly Germany Although there have been small scale attempts at the development of techniques for predicting success in science prior to the second World War, it was only during and after it that national efforts were organized to locate scientific talent and to encourage its training With such ability at a premium it seems likely that its selection will receive even more attention in the future than medicine has in the past and than psychology is receiving at the time of writing

The Stanford (or Zyve) Scientific Aptitude Test (Stanford University Press, 1929) is probably the first published attempt to develop a measure of scientific aptitude, but little work has been done with it since either by its author or by others despite its continued use The test attempts to measure the components of scientific aptitude, science being defined as

organized knowledge based on experiment and observation The test therefore consists of eleven parts, designed to measure experimental bent by expressions of preference for experimental as opposed to bibliographical or other methods of obtaining information, clarity of definitions, suspended versus snap judgment as manifested in ability to state that answers to problems are not available, reasoning concerning physical problems (in four parts differing in content), caution and thoroughness as demonstrated in the solution of apparently easy problems, ability to select and arrange experimental data for the solution of a problem, comprehension of scientific reading matter, and perception of complex spatial detail The items were developed and checked with the aid of established scientists, and were validated against grades in scientific courses The correlation with intelligence tests, according to the manual, was found to be 51 with college students The correlation with the grades of science students was 50, in contrast with that of 27 for the Thorndike Intelligence Examination, the correlations with grades of non-scientific students were respectively 02 and from 38 to 53, which strongly suggests that the test does measure intellectual factors which are important to success in scientific but not in literary endeavor

The Stanford test was administered by Benton and Perry (75) to 43 students (30 science majors, 13 others) at the College of the City of New York They found correlations of 30 and 37 between this test and four-year grades, while intelligence as measured by the A C E Psychological Examination had a validity of 31 with total grades, 27 with science grades, and 41 with non-science grades The intercorrelation of the two tests was 45 Studies of this test have been so few and are so inconclusive that it is difficult to judge its validity, especially when the attenuation of validities usually noted in studies made after the original authors' are kept in mind

The Science Talent Search administered by Science Service and financed by the Westinghouse Electric Corporation is a project in which one might expect to find a battery of tests for the selection of potential scientists being developed The selection procedure consists of a series of five hurdles a Science Aptitude Examination, high school grades, a recommendation by teachers, an essay on a scientific topic, and psychological and psychiatric interviews (235) The Science Aptitude Test first used was a reading test of scientific subject matter, but in later years what amounted to a battery of tests was utilized A variety of types of items were used, including both scientific vocabulary and Bennett-type

mechanical comprehension pictures, scores were independent of amount of mathematics and science studied, but validity data have not as yet been made available, making evaluation of the procedure impossible at present

"Scientific aptitude" being presumably largely an intellectual matter, it seems likely that batteries of tests for the selection of promising scientists will stress such factors as reasoning, spatial visualization, and number ability, scientific vocabulary and mechanical comprehension are two less pure aptitudes which should also be significant, and inventoried interest may prove to have value for completion and occupational utilization of training if not for quality of work done It seems strange that work has not been done with such a battery

Teachers Tests of aptitude for teaching have been experimented with by a number of individuals and schools of education, in attempts to improve the selection of students of education The New York State Department of Education and the Psychological Service Center of George Washington University, are among the institutions which have published custom-built tests of so-called teaching aptitude Other institutions such as the University of Wisconsin and the University of California at Los Angeles have worked with batteries of standard tests in attempting to develop sound selection procedures Tests for the evaluation of preparedness for teaching have been prepared by the Educational Testing Service as the National Teacher Examinations (28 9), administered annually to candidates for teaching positions who wish to have an objective record of their mastery of subject matter made available to possible employers (263)

The Coxe-Orleans Prognosis Test of Teaching Ability (World Book Co, 1930) is a good example of custom-built tests of aptitude for teaching It consists of five subtests general information, knowledge of teaching methods and practices, ability to learn the type of material included in professional texts, comprehension of educational reading matter, and judgment in handling educational problems Validation of this instrument has been in terms of success in teacher training, but the data are not very helpful because they consist of correlations between the prognostic test and a comprehensive achievement test at the end of the first year of training These coefficients range from 53 to 84 as cited in the manual, but in view of the highly academic nature and similar content of both tests the evidence is not convincing

It has apparently not been validated against criteria of success on the

job In view of the difficulties commonly encountered in establishing criteria of success in teaching this is perhaps understandable The validity coefficients will undoubtedly be much lower than those reported in the manual, since teaching is less exclusively dependent upon intellectual ability than is learning about teaching

Seagoe's studies (687,688,689) are a good illustration of work with standard tests in the selection of students in schools of education She administered the American Council Psychological Examination, Cooperative General Culture Test, Meier Art Judgment Test, Seashore Tests of Musical Talents, Strong Vocational Interest Blank, Allport-Vernon Study of Values, Bell Adjustment Inventory, Bernreuter Personality Inventory, and Humm-Wadsworth Temperament Test, to 125 students of education Ratings of success in two practice-teaching assignments were obtained for 31 of these students, and were correlated with the test scores (688) No significant relationships were found between the tests of intelligence, special aptitudes, achievement, interest, or values and the ratings of success in practice teaching, relationships between personality inventory scores and ratings were significant, those for the Bell keys being — 10 (total adjustment) and that for the Bernreuter Self-Confidence scale being — 38 Twenty-five of these students were followed up after two years of teaching in the field, using rank in the faculty as judged by the school administrator as criterion (689), the Bell and Bernreuter were again found to have some validity, as did ratings by critic teachers, grade-point ratio had none

The numbers in Seagoe's studies, as in other studies of the same type are small and criteria of success need to be improved, before objective selection procedures can be considered adequate in this field But as long as teaching remains an underpaid occupation with too few applicants for available positions there is not likely to be much pressure for the development of better selection methods, at least in most training institutions

The National Teacher Examinations (Educational Testing Service, annually since 1939) provide school systems and graduate schools of education which can afford to be selective with a standard battery of tests for the evaluation of teachers' mastery of subject matter, reasoning, and judgment These are, obviously, only intellectual aspects of ability to teach, and do not include interest in children, emotional stability, and other factors which are generally believed to be important to teaching success But Flanagan (263) found that scores on this battery of tests

had a correlation of 51 with ratings of 49 teachers in 22 school systems made by two supervisors and five students in each case, which indicates that the tests have value in selecting good teachers despite the fact that they do not measure everything that is to be considered As Flanagan points out, other characteristics must be appraised by means of interviews, ratings, and recommendations in the absence of more objective methods

CHAPTER XV

STANDARD BATTERIES WITH NORMS FOR SPECIFIC OCCUPATIONS

THE characteristics, advantages, and disadvantages of standard batteries consisting of generalized items which can be validated and weighted as tests rather than as items, and for which norms can be developed for a great variety of occupations, have been discussed at the beginning of the preceding chapter In this chapter, therefore, it is necessary only to describe and discuss the two such batteries which are currently coming into use the *General Aptitude Test Battery* of the United States Employment Service, and the *Differential Aptitude Tests* of the Psychological Corporation It might be added parenthetically that other such batteries are being published, notably by Guilford (320) and by the American Institute for Research, but that the task of obtaining occupational norms is so great that only well-financed organizations can ethically undertake it The day of the publication of isolated tests of single aptitudes will no doubt soon be past

The General Aptitude Test Battery (United States Employment Service, 1947)

This battery is the product of more than ten years of research in worker characteristics and test development by the Occupational Analysis Division of the United States Employment Service, described most completely in two journal articles by Shartle, Dvorak, Heinz, and others (735,225) This comprehensive program of research in vocational aptitudes was itself the outgrowth, insofar as principles and technical matters are concerned, of the Employment Stabilization Research Institute of the University of Minnesota, the work of which has been frequently encountered throughout this book With such a long and fruitful history behind it, it is natural to expect that this battery should prove a landmark in the history of the appraisal of vocational promise Dvorak's

description of it (224,225) encourages such expectations She states "With the General Aptitude Test Battery, however, it is possible to obtain information about an individual's aptitude for several thousand occupations in little more than two hours of testing" It is explained in another paragraph that this is done by means of norms for 20 fields of work, representing nearly 2000 occupations grouped as in Part IV of the *Dictionary of Occupational Titles* (888) but, in this case, on the basis of similar minimum amounts of the same combination of aptitudes

This represents a very real accomplishment, as is evidenced by the meagerness of the occupational norms which discussions of the majority of tests in this book have revealed Unfortunately, Dvorak's two identical articles (224,225) are lacking in many of the details which are necessary for judging the adequacy of the norms, validities, and other basic data concerning tests, data which are now rather routinely reported by professionally competent and ethical test constructors and publishers As the tests are not available for use outside of the United States Employment Service and cooperating public schools this is not a matter of practical urgency to counselors as users of tests, but it is a matter of great importance to personnel men, to the federal and state governments, and to the profession as a whole that the tests used by Employment Service Counselors be not only adequate but demonstrably so In the discussion which follows, based on Dvorak's articles, on the tests themselves, and on the training manuals and directions which accompany them, some of the important unknowns will be brought out, it is to be hoped that subsequent publications will provide the needed information

Applicability The General Aptitude Test Battery was developed for use with adult employment applicants, including older adolescents recently out of school, who are in need of vocational counseling in connection with registration at the offices of the federal-state Employment Service It is to be used when other evidence concerning aptitudes is unsatisfactory, when other important abilities are suspected, when the applicant has difficulty choosing among several seemingly suitable fields, and when the applicant needs a better understanding of his vocational strengths and weaknesses No data are available concerning the differences in the performances of adolescents, young adults, and older persons, they would be desirable as a guide to interpreting the scores of recent high school graduates

Contents The battery consists of 15 tests, the scores of which are combined to yield scores for 10 factors The paper-and-pencil tests are

printed in two booklets totaling 70 pages, the apparatus tests consist of a rectangular manual dexterity box or pegboard and a small rectangular board for the finger dexterity test The subtests in the booklets are as follows Tool Matching, a test for perception of similarities and differences in the black and white shading of simple pictures of familiar tools, Name Comparison, resembling the Minnesota Clerical (Names) Test, H-Marking, somewhat like the MacQuarrie (Pursuit) Test, Computation, consisting of addition, subtraction, etc , Two-Dimensional Space, resembling the Revised Minnesota Paper Form Board, Speed, like the Dotting Test of the MacQuarrie, Three-Dimensional Space, a metal or paper-folding test, Arithmetic Reasoning, verbally expressed arithmetic problems, Vocabulary, a same-opposites test, Mark-Making, a manually more complex dotting test, and Form Matching, like the analogies tests of the A C E Psychological Examination The Pegboard yields two scores, one for placing and one for turning, as in the Minnesota Manual Dexterity Test, but the pegs are smaller than the disks of the latter test, and both hands are used in placing The Finger Dexterity Board is administered for both assembly and disassembly The USES policy appears to have been to construct items as much as possible like those of earlier standard tests which had proved valid

Administration and Scoring Administration of the General Aptitude Test Battery requires about two and one-quarter hours The two booklets of paper-and-pencil tests are designed for group testing, this is true also of the apparatus tests, which are so constructed that in taking one part of the test the examinee automatically sets them up for the next test Answers to paper-and-pencil tests are recorded in the test booklets, which makes testing somewhat more expensive than it would be with special answer sheets, but the additional expense may be warranted by the greater ease of administration to a heterogeneous population Stencils are provided for scoring, which is objective and simple Raw scores for each part are changed to "converted scores" by means of a conversion table, these are summated by groups to provide "aptitude scores" for each of the 10 factors measured by the 15 tests These are standard scores, with a mean of 100 and a standard deviation of 20

The 10 aptitude scores obtained from the 15 tests are described as follows

G—*Intelligence* general learning ability, ability to grasp instructions and underlying principles It is often referred to as scholastic aptitude

V—*Verbal Aptitude* ability to understand the meaning of words

and paragraphs, to grasp concepts presented in verbal form, and to present ideas clearly

N—Numerical Aptitude ability to perform arithmetic operations quickly and accurately

S—Spatial Aptitude ability to visualize objects in space and to understand the relationships between plane and solid forms

P—Form Perception ability to perceive pertinent detail in objects or in graphic material, to make visual comparisons and discriminations in shapes and shadings

Q—Clerical Perception ability to perceive pertinent detail in verbal or numerical material, to observe differences in copy, tables, lists, etc It might also be called proofreading

A—Aiming or Eye-Hand Co-ordination ability to co-ordinate hand movements with judgments made visually

T—Motor Speed ability to make hand movements, such as tapping, rapidly

F—Finger Dexterity ability to move the fingers and to manipulate small objects rapidly and accurately

M—Manual Dexterity ability to move the hands easily and skillfully, a grosser type of movement than finger dexterity, involving the arms and even the body to a greater extent

It can be seen from the above that the General Aptitude Test Battery measures most of the aptitudes which have so far been isolated There is no measure of mechanical comprehension, but we have seen that this is not a factorially pure aptitude, but rather a composite of aptitude and experience, of which spatial comprehension is the major component Artistic judgment and the musical capacities are not tapped, but they are of very specialized significance and perhaps wisely omitted from a general aptitude battery Interests and personality are not assessed, but these are not aptitudes The GATB therefore includes all of the aptitudes discussed in this book, all of those isolated in earlier factor analyses of abilities except memory (if Thurstone's Reasoning and Induction factors may be considered subsumed in G), and some newly isolated factors

Norms No mention is made, either in Dvorak's paper or in the manuals published for use of the Employment Service, of the number of persons in each occupation or field for which norms are provided These may be large and representative both as to fields and as to parts of the country, but evidence on the matter has not been presented either to the public or to the staff members of the Employment Service who use

the tests and should have data as to their scientific basis There is every reason to assume that an agency with the resources of the USES, and persons with the test construction experience of Shartle and Dvorak, would do a workmanlike job of developing norms, on the other hand, the admittedly preliminary work described in Stead and Shartle (750) involves numbers which are smaller than one would like, and the occupational ability patterns developed for use in selection for specific jobs (a series of batteries quite distinct from the GATB) were, at the time of writing, based on so few cases that they were used tentatively and with extreme caution, and then only by well-qualified examiners It is to be hoped that data on the numbers involved in each of the 20 fields and 2000 occupations will be made available

The occupational-field norms are utilized to establish cut-off scores for each aptitude which plays a significant part in each field Thus Occupational Aptitude Pattern No 1 has a cut-off standard score of 130 for G, general intelligence, and of 130 for Verbal Ability, this pattern is for a field which includes literary work (D O T code O X3), creative writing (O-X3 1), and copy writing and journalism (O-X3 5), the field might perhaps be called the Literary Field, although the fields have not been officially named because of the fact that the development of more tests and the establishment of patterns for more occupations may change their apparent nature For example, what seems to be an electrical assembly field will probably include other types of small, but not fine, routine technical assembly work as other occupations are studied The cut-off score for a given aptitude for a given occupation is that below which one-third of the occupational group in question were found to fall The publications give no reason for the selection of this point rather than the quartile or some other figure Cut-off scores in selection programs are based on the percentage of satisfactory workers which would have been accepted, and the percentage of unsatisfactory workers which would have been rejected, on the basis of that cut-off point, but in guidance the establishment of such cut-offs is extremely difficult because of varying criteria The use of production and worksample criteria in the preliminary USES studies (750) suggests that this may have been done for the various selection batteries, if it was done for the GATB it should be described If, on the other hand, the cut-off score was established at the 33rd percentile merely as a point distinguishing a "less able" from a "more able" group of workers, so labeled on the basis of tests whose validity was believed to have been sufficiently demonstrated in previous

studies by other psychologists, this also should be made clear Such a cut-off score is useful, but being below it is less prognostic of failure than when the cut-off score is a point below which few succeed

Standardization and Initial Validation No sequential picture of the development of the General Aptitude Test Battery has as yet been published But it has no mean history, and an integrated account of the work of which it was a part would have considerable value The genesis of the idea was in the Minnesota Employment Stabilization Research Institute, written up by Paterson and Darley (589) and by Dvorak (223), early work by the USES is described by Stead, Shartle, and others (750), but this work was still partly with published tests and did not concern the General Aptitude Test Battery a factor analysis of these and other tests was published by the Staff of the Occupational Analysis Division in 1945 (735), and the Dvorak articles (224,225) describe the battery and the *procedure* of standardization and validation without giving any of the *results* Even the procedural material is general in nature As described by Dvorak the standardization procedure began with job analysis, to identify the job and define the sample population Persons were then included in the sample if they were performing the same type of work, had passed the learning stage, and were rated satisfactory by their supervisors Care was taken to make the samples all-inclusive or representative Although she thus describes the sampling procedure, Dvorak says nothing about the construction of the tests prior to standardization testing Neither does she describe the validation or norming processes, beyond stating that the cut-off scores are placed at the point which eliminates the lowest third of the occupational group

In the paper in which Dvorak collaborated with other staff members (735) somewhat more data are given in connection with the USES's factor analysis study In this report nothing is said specifically about the GATB but it is evident from the discussion and from the tests listed that it was included, along with 44 other tests Based on this total of 59 different tests, administered in various combinations to groups of from 99 to 1079 persons, or a total of 2156 individuals, in 13 different communities scattered across the country, this is one of the most thorough factor analyses of aptitude tests which has been made It is therefore regrettable that here, too, the presentation is rather general and omits much of the detail which the careful reader of the literature and the conscientious and insightful user of tests needs

Despite these limitations the report is helpful It gives some idea of

the empirical justification for using the tests which are in the battery, and especially for combining them to yield factors or aptitude scores This is fortunate, as it is in this respect more than in any other (except its occupational norms) that this battery differs from the Psychological Corporation's Differential Aptitude Test Battery There is, for example, the justification for grouping three of the GATB tests (Three-Dimensional Space, Arithmetic Reasoning, and Vocabulary) to yield a score for general intelligence One's first reaction might be to assume that this was merely a catering to the layman's desire to think in terms of "intelligence" because tests have yielded such scores for a generation, on the contrary, the report of the factor analysis makes it clear that it was a step made necessary by the evidence As the authors state "it appears to have some of the properties of Spearman's G (sic), but the two-factor theory has no place for group factors like *V*, *N*, or *S* (which also were isolated) On the other hand, this factor has a wider significance and is more persistent than either Thurstone's *R* or *I* It appears to possess many of the properties that teachers, test examiners, and clinical psychologists would attribute to 'intelligence' this factor has been designated, noncommittally, as Factor O" In the manuals it is designated as *G*, and is uncompromisingly called intelligence It is interesting that this finding of a general intelligence factor was accomplished, not with Spearman's two-factor statistical procedures, but with the use of Thurstone's centroid method of factor analysis, which has not on other occasions revealed a general factor Furthermore, the sample was one of young adults, aged 17 to 39, rather than one of children in whom maturation rates would tend to produce a seemingly general factor

In view of the fact that studies such as these were part of the process of developing the battery, it seems legitimate to assume that the procedures of constructing the actual tests in the battery were well conceived and carried out It is to be hoped, however, that more of the details of this procedure, of the sampling procedure described above, and of the validation procedure will be published

The only available evidence of validity lies in the cut-off scores for the various occupational groups, and this is only implicit evidence not analyzed or reported as such, the published material says nothing about validation The fact that the Verbal Aptitude cut-off (standard) score of "literary" workers is 130, while that of "copy" workers (the terms in quotations are the writer's) is 100, and that the Form Perception cut-off

score of "technical assembly" workers is 100 while that of "routine assembly" workers is 85, is evidence of occupational differentiation and therefore of validity The 20 occupational fields, tentatively named for convenience's sake by the present writer, are listed below, together with the codes and titles from Part IV of the *Dictionary of Occupational Titles* (888), the aptitudes required and representative occupations They provide evidence not only of the validity of the battery (its ability to differentiate between occupations) but also of its significance for the classification of occupations

1 *Literary occupations*, 0-X3, require a high degree of general intelligence and verbal ability, they include creative writing, translating, copy writing, and journalism

2 *Computational work*, 0-X7 1, embraces the accounting occupations, it is engaged in by persons with a high degree of intellectual and numerical ability

3 *Engineering occupations*, 0-X7 4, include at least some of the engineering fields, the aptitudes required being intellectual, numerical, and spatial in high degree and form perception in a moderately high degree

4 *Technical-mechanical work*, 4-X2 010 and 4 X2 100, requires average amounts of intelligence, number ability, and spatial ability, and a fair degree of finger dexterity The field includes machine-shop and all-around mechanical repair occupations

5 *Record work*, 1-X1, 1-X2 0, involves average general intelligence, moderately high numerical ability, and average clerical perception Included are routine computing and general recording

6 *Artistic design occupations*, 0-X1, are characterized by moderately high intelligence, average spatial ability, and moderately high form perception The field includes artistic drawing and arranging

7 *Technical-electrical work*, 4-X6 18, requires a fair degree of intelligence, together with average spatial ability, form-perception, and finger dexterity It includes electrical wiring and radio repair

8 *Copy work*, 1-X2 2 and 3, 4-X6 56, is performed by persons, the majority of whom have average or better verbal ability and clerical perception, and fair motor speed and finger dexterity Occupations are both clerical (typist, stenographer) and skilled (typesetter, hand composer)

9 *Mechanical work*, 1-X2 103 and 4-X2 104, is characterized by average or better numerical and spatial abilities, and fair form perception and

manual dexterity The occupations known to be included are combustion engine and aircraft equipment repairman

10 *Industrial design,* 0-X7 7, involves moderately high numerical, spatial, and form-perception abilities, and average or better aiming or eye-hand co-ordination Typical occupations are various kinds of drafting

11 *Routine recording,* 1-X2 8, involves average clerical perception and fair numerical ability, and includes not only routine record-keeping jobs but also equipment and material checking

12 *Business machine operation,* 1-X 1, 1-X 2, differs from record work (#5) in that it requires less general intelligence, only average or better numerical ability, and, in addition, average or better motor speed and fair finger dexterity, it also requires average or better clerical perception The occupations included have the same Part IV DOT classification, which suggests that the latter does not make sufficiently refined distinctions in this area category No 5 is more mental, No 12 more mechanical

13 *Structural work,* 4-X6 2 is characterized by fair numerical, spatial, and manual abilities it includes not only structural work with heavy metal, but also plumbing and carpentry

14 *Technical assembly,* 4 X6 3, requires average or better form perception and finger dexterity, and fair spatial and aiming abilities Types of assembly are electrical units, mechanical units, and optical units, including repair

15 *Shaping work,* 4-X6 3, involves manual rather than finger dexterity, demands no special facility in eye-hand co-ordination, is otherwise like the technical assembly field, it includes grinding and tool dressing

16 *Visual inspection,* 4 X6 38 and 6-X2 38, requires fair form perception, whether for close or simple visual inspection

17 *Routine assembly,* 6-X4 30, is characterized by average or better eye-hand co-ordination and finger dexterity, and form perception Simple electrical unit assembly jobs are the only type so far included, but no doubt certain nonelectrical jobs will be found in the same category

18 This heterogeneous category cannot at present be named The common characteristics are average form perception and fair manual dexterity, the occupations include such metal trades as roller and extruder, range through various stone setting jobs, and also include visual

inspection jobs in metal, leather, lumber, meat-packing, and other industries

19 *Classifying clerical work*, 1-X4, requires average or better clerical perception and motor speed, includes classifying jobs such as file and mail clerks and directory compilers, and other clerical workers such as office boys and sorters

20 *Machine operation*, 6-X4 4, involves fair amounts of eye-hand coordination, motor speed, and finger and manual dexterity Occupations include a great variety of machine operating and tending jobs, from machine sewing through metal polishing, wood sanding, and printing press feeding and catching, to pipe bending

Reliability No reliability data have yet been published In view of the types of items, the amount of work done with the battery, and the qualifications of those supervising it, it seems hardly likely that they are below 85 But this should be made explicit

Validity As the battery was tentatively put into use by the Employment Service in the spring of 1947, and is restricted to that organization, no studies of its validity as such have as yet been published in the literature In view of the long range program which produced it, it seems very likely that the General Aptitude Test Battery will prove to have considerable validity As its widespread use in the Employment Service makes possible the rapid collection of data concerning large numbers of people entering many different occupations, objective data concerning its value in counseling as opposed to discriminating between persons employed in various fields should be relatively easy to collect

Use of the General Aptitude Test Battery in Counseling and Selections As this battery of tests is designed only for Employment Service use, at least provisionally there is in one sense no need to discuss their use in this treatise A few points, however, are worth noting for their general significance One is that the battery, although designed for counseling and standardized with that in mind, could equally well be used for selection purposes, it is composed of relatively pure tests, factorially speaking, gives a variety of scores which seem to be of occupational significance, and could well be validated against local criteria in a student or employee selection program Secondly, as the tests have all been standardized on the same population, that upon which the standard scores are based, have norms for a variety of occupations expressed in the cut-off scores, and are to be administered to additional occupational groups for

norming purposes, the battery is potentially the most useful instrument of individual diagnosis which has been developed It should almost certainly prove extremely valuable in colleges, guidance centers and employment services, in dealing with young adults If the normative data were extended downward by the establishing of age and grade norms (a much easier task than extending grade norms upward), and by ascertaining the effect of maturation and experience on the test scores, it could become an extremely useful instrument in the schools It is for these reasons that the battery has been described in so much detail in this text, even though most users of this book do not now have access to it It is to be hoped that, tentative and incomplete though it is, it has established a pattern which will be further developed and become the pattern for the future

The Differential Aptitude Tests (Psychological Corporation, 1947)

This battery of tests was developed by Bennett, Seashore, and Wesman, in response to widespread feeling among vocational psychologists and counselors that a major defect in current testing programs is the lack of a uniform baseline for the various tests which are used with a given student or client (Manual A-3) We have seen, for example, that the Revised Minnesota Paper Form Board has norms which are based on differing groups in a few localities, and that the Bennett Mechanical Comprehension Test has a totally different and equally limited base A student may be at the 65th percentile when compared to liberal arts college freshmen on one test, and at the 55th on another, but actually have more ability of the type measured by the second test the seemingly lower score may be due to differences in the normative groups It is only when the tests in a battery have been standardized on strictly comparable groups, if not the same group, that one can effectively study aptitude or interpret differences *within* individuals

Other needs also contributed to the development of this battery One was the improvement of statistical procedures which made possible the construction of tests which effectively measure narrower aspects of ability than general intelligence We have already seen the development of quantitative and linguistic scores for the A C E Psychological Examination, the Wechsler-Bellevue Scale, and other modern substitutes for the undifferentiated tests of the times of Binet and Otis, and the further development of factor scores by Thurstone in the Primary Mental Abilities Tests Still another was the time factor, for it is important that a

comprehensive battery be administrable in a reasonably brief period if educational and occupational norms are to be obtained for all tests from the same subjects It is a sign of the times that both the United States Employment Service and the Psychological Corporation have moved simultaneously to meet these needs. The American Institute for Research is preparing an integrated battery of its own, also for use in guidance, Guilford (320) has released a similar battery, and it seems likely that other test publishers will be forced either to follow suit in due course (an expensive process) or to confine their energies to specialized fields such as achievement, special talents (artistic, musical, manual), interest, and personality And even some of these will probably be removed from the list as standard batteries are improved, for there is evidence which suggests that paper-and-pencil tests of manual dexterities and interest inventories (and perhaps tests) can be developed which will be much more valuable if used as parts of an integrated battery

Applicability The battery was designed for use with high school students, including 8th grade boys and girls Items were devised for and retained on the basis of their suitability for this age and ability range, and the time limits and norms are based upon the performance of samples of high school populations They may therefore be considered extremely effective at these levels No attempt was made to make the tests applicable to college students or adults, although use in personnel selection was envisaged (Manual A-1, A-5) and the items may well be suitable, but the fact that the grade norms increase annually from grade 8 through 12 shows that special *norms* would be necessary As age norms have not yet been provided, and no analysis has been made of the effects of progressive elimination in high school on the sample, it is still impossible to draw any conclusions concerning the development of these abilities from the preliminary work with these tests seniors may make higher average scores because they have lived and studied one year longer than juniors, or because they have lost their less able classmates by the wayside Be this as it may, the development of college or adult norms has one possible drawback in the ceiling of the tests having been designed for high school students, they might not permit the most able college students and adults to show the full extent of their abilities

Content. The Differential Aptitude Tests consist of eight tests designed to measure eight different abilities Some of the abilities are aptitudes in the stricter sense of the term (Verbal Reasoning, Numerical Ability, Space Relations, and perhaps Abstract Reasoning), others are factorially less

pure (Clerical Speed and Accuracy and Mechanical Reasoning) but can be treated as aptitudes, while still others are proficiencies (Language Usage Spelling and Sentences) The last-named are, however, sufficiently basic forms of achievement to be used effectively as indices of promise Because of the excellent descriptions in the manual, the following paragraphs are in part abstracts of the manual

The Verbal Reasoning Test attempts to measure ability to generalize, to think with words à la Thurstone (V) It consists of verbal analogies, in which the first member of the first pair and the second member of the second pair have been omitted from the stem and must be selected from two sets of items with four choices each, thus _____ is to x as y is to _____ Analogies were used because they have proved to be one of the best types of reasoning test items, and the form chosen is highly reliable, versatile, and lends itself to complexity without resort to esoteric terms Because of this latter fact, the vocabulary is relatively simple, the content familiar, and complexity is a function of the reasoning processes involved

The Numerical Ability Test is designed to measure understanding of numerical relationships and facility in handling numerical concepts, another of Thurstone's factors (N) As the manual points out, the items are cast in the form usually referred to as "arithmetic computation" rather than "arithmetic reasoning," the reason given is that language problems are thus avoided, and that complexity was attained by the numerical relationships and the processes to be used in the problems

The Abstract Reasoning Test attempts to measure reasoning without the use of words (Thurstone's R) Problems are of a spatial type made familiar by the A C E Psychological Examination, and require finding the principle underlying a series of changing geometric figures

The Space Relations Test (Thurstone's S) is the most ingenious in the series, although embodying familiar principles These are ability to visualize a constructed object from a pattern (structural visualization in three dimensions), and ability mentally to manipulate a form in order to judge its appearance after rotation in various ways By combining these two principles in items which require the mental folding of cut or partly shaded patterns a test of spatial visualization has been developed which promises to be superior to any so far developed

The Mechanical Reasoning Test is another form of the familiar Bennett Mechanical Comprehension Test The mechanical principles are illustrated with pictures of familiar objects, but care was taken to avoid textbook illustrations

The Clerical Speed and Accuracy Test is designed to measure speed of response to numerical and alphabetical symbols Although presumably a substitute for the Minnesota Clerical Test, it differs considerably from the latter in its mechanics, and also seems to differ in its factorial composition, for letter and number-letter combinations are substituted for the names used in the older test The examinee finds the underlined combination in each row of a block of symbols, then marks the same combination (differently placed) in the same row of the same block on the answer sheet Intelligence plays a less important part in this task than in the Minnesota Names Test

The Language Usage Test contains two parts, *Spelling* and *Sentences* In the former, each word is marked as spelled either right or wrong, in the latter each sentence is divided into parts, to be marked according to their correctness The types are familiar, the items chosen by established scientific procedures

An attempt was made, in drawing and printing the items in these tests, to make them sufficiently large and clear so that visual acuity would play no part Inspection of the items does suggest that they are free from some of the defects which can be noted in certain other tests involving mechanical objects, geometric figures, and other drawings in which details might be obscure or irrelevant differences slight and confusing

Although the test authors point out that the Differential Aptitude Tests were designed, not to measure all known and mensurable aptitudes, but rather to measure a number of important variables which have meaning for vocational counseling and selection and which can be assessed in a reasonable period of time, one cannot help but check the aptitudes tapped by these tests against those assessed by the USFS General Aptitude Test Battery and isolated by various factor analysis studies (735,839) The Verbal, Numerical, Spatial, Abstract Reasoning, and Clerical Speed and Accuracy Tests clearly correspond to the verbal, numerical, spatial, reasoning and perceptual factors isolated by Thurstone and by Shartle and associates, and measured by the General Aptitude Test Battery The Mechanical Reasoning Test has no counterpart in the GATB, presumably because it taps a composite of factors rather than one factor, neither do the Language Usage Tests, which are achievement measures On the other hand, Thurstone isolated a memory factor (not reliably measured) and the GATB provides measures of eye-hand co-ordination, motor speed, finger dexterity, manual dexterity (the last two require apparatus tests), and distinguishes between form and clerical perception

This suggests that Bennett, Seashore, and Wesman have gone further in the direction of measuring what counselors look for and what has proved to have validity than did Shartle, Dvorak, and associates, and, conversely, that the latter have attempted more consistently to make useful the findings of the factor theorists Given adequate norms and validation data, the USES policy may prove wiser in the long run, until then, the Psychological Corporation's policy of providing measures of types which have known occupational validities may be sounder

Administration and Scoring The eight tests are printed in seven booklets (the two Language Usage Tests are in one booklet), making possible administration of any of the tests in any order desired Time limits are such that any test can be given in one class period, they vary from six minutes to 35 minutes Total testing time is three hours and six minutes The manual recommends that the tests be given in an order which will hold interest and avoid monotony, and suggests two arrangements, one of three and one of two testing sessions, which are not quite identical This raises the interesting question of the possible effect on profile scores of testing some students with one sequence, some with another, and of testing some students in few sessions, some in several The test authors do not mention this problem, which may not be an important one, but until it is demonstrated to have no effect it is probably wise to adopt a set sequence and spacing of tests and to follow it rigidly, thereby making all local scores comparable with each other if not actually with the national norms (it is not clear just what sequence and spacing were used in gathering norms, nor even that the procedure was standardized in this respect) Answers are recorded on IBM answer sheets, making possible either hand-stencil or machine scoring The manual contains unusually complete suggestions for efficient test administration and scoring, from advance arrangements to a summary table of scoring information, incorporating the best experience of the large-scale testing programs of recent years

Norms Norms are available for each grade from 8th through 12th, and for each sex, for both forms of the test They permit the conversion of raw scores into percentiles, which were adopted instead of standard scores because of their current widespread use, the profiles permit conversion into approximate standard scores, and such a system is to be made available in due course because of that system's more accurate representation of individual differences The students on whom the tests were standardized were enrolled in schools scattered throughout the Eastern

and Midwestern states, Western and Southern norms are in preparation, industrial and business norms will be provided routinely to manual owners as research projects make them available The Eastern and Midwestern norms are based on 30 school systems, ranging from Yorktown Heights (a small northern Westchester County suburb of New York City) and Gloucester (a small Massachusetts fishing and resort city) to Ann Arbor (the University of Michigan's college town) and St Paul (Minnesota's industrial city) In some communities all pupils in all five grades were tested, in others, representative samples (as judged by the local research director) were tested Form A was standardized on the largest groups, these range in numbers from 482 for the 12th grade boys to 1561 for the 9th grade boys, and from 578 12th grade girls to 1642 9th grade girls For a preliminary standardization such regional coverage and numbers are almost unique, they appear to be such as to make possible the use of the tests at once in Eastern and Midwestern communities Judging by the results of other tests these norms will be somewhat high for Southern states, and somewhat low for the West Coast, but other regional norms to be published will soon, no doubt, be available It is to be hoped that curricular norms will become available, and that college freshmen norms, based on homogeneous and well-described types of colleges, will also be compiled and published

Standardization and Initial Validation What has previously been said concerning the content, the development of norms for this battery of tests, and data in the subsequent paragraphs on its reliability, conveys an adequate idea of the work which was done in standardizing these tests The types of items to be included were decided upon on the basis of factor analysis and validation studies carried out by other psychologists with other tests The test items were tried out in preliminary studies and the tests were administered for standardization purposes only when they seemed administrable Care was taken to obtain large samples of students at each appropriate grade level and in representative communities The reliability of each test was computed and, with one limited exception, found adequate for individual diagnosis (see below) Finally, the intercorrelations of the tests were obtained These latter ranged, for Battery A (boys), from 06 (Mechanical Reasoning and Clerical Speed and Accuracy) to 62 (Verbal Reasoning and Language Usage Sentences), data for girls, and for Battery B, were approximately the same The median intercorrelation for Form A tests is 425 These intercorrelations are not much higher than those of the Primary Mental Abilities Tests, after allowance

is made for the achievement (Language Usage) and composite (Mechanical Reasoning) tests, for which the two highest correlations with other tests were obtained Knowledge of the educational and vocational predictive value of similar tests which have already been discussed (the Primary Mental Abilities Tests, A C E Psychological Examination, Minnesota Paper Form Board, Bennett Mechanical Comprehension Test, Minnesota Clerical Test), combined with the proved reliability and relative independence of these tests suggests that studies using external criteria should demonstrate considerable validity for these tests

Reliability Particular care was taken, in establishing the reliability of the Differential Aptitude Tests, to avoid the common defect of tests with part scores, that is, reliability of the total score but insufficient reliability of the part scores for individual diagnosis Homogeneous groups were used, to avoid the spuriously high coefficients which are yielded by heterogeneous groups Split-half reliabilities were computed for all but the Clerical Speed and Accuracy Test, for which as a speed test, that technique is not suited, instead, alternate-form reliability was ascertained The Form A reliability coefficients for 960 boys range from 85 (Mechanical Reasoning) to 93 (Space Relations), for 1064 girls they ranged from 71 (Mechanical Reasoning, a type of test which generally has little value for girls) or 86 (Numerical Ability the second lowest for girls) to 92 (Language Usage Spelling) For boys, then, all of the tests in Battery A have quite adequate reliability, for girls, all those which are likely to be useful have equal reliability Data for Battery B are about the same, this form of the Mechanical Reasoning Test having been revised and improved

Validity The Differential Aptitude Tests being recently published, there has been little time for the carrying out of studies of their validity in relation to external criteria The authors felt that the known significance of the abilities measured, combined with the internal evidence of validity, was sufficient to justify making the test available at this stage of development (Manual E 2) As they have committed themselves to an extensive program for the validation of the battery against educational and occupational criteria (Manual E-1), and as the early publication of a reliable test has been demonstrated to speed up its further validation by other investigators (e g , Kuder's work, Ch 18), this would seem to be quite justifiable A supplement to the manual now includes a large number of validity coefficients, based on the high school grades of norm groups

Use of the Differential Aptitude Tests in Counseling and Selection The preliminary evidence concerning the development and standardization of the DAT battery suggests that these tests measure a number of variables which have frequently been found to have vocational significance For an understanding of the development and vocational significance of the traits measured by these tests, the chapters dealing with similar specific tests should be studied

In schools and colleges, when clinical counseling is to be done, that is, when the objective is the study of a counselee in terms of his psychological make-up and its general educational and vocational implications, the battery should prove useful When, however, comparisons need to be made with pre-occupational or occupational groups, the lack of occupationally differential norms renders this battery temporarily useless As there is every reason for believing that curricular and occupational norms will be developed, counselors in schools and colleges may want to use this battery for clinical counseling, developing their own curricular and vocational norms as part of their follow-up work

Guidance and employment centers which habitually carry on normative studies may also find it worth their while to use this battery of tests in clinical counseling, supplementing it with others which have occupational norms when such data are really needed Other tests, such as those of manual dexterity, may be needed in any case to round out the picture, together with personal data obtained in interviews If the battery is used, it should be only with a definite co ordinate research program in mind This can be materially aided when the center works co-operatively with business and industry in employee selection programs

In business and industry, even more than in guidance work involving clinical counseling, the gathering of local norms and validation against local criteria should precede the use of the results of these tests for selection purposes Validation for selection is so much easier than validation for counseling, and the accuracy of predictions is improved by so much greater a degree, that to adopt any other policy is to be guilty of gross negligence

CHAPTER XVI

THE NATURE OF INTERESTS

INTERESTS have probably received more attention from vocational psychologists during the past generation than any other single type of human characteristic, including intelligence, aptitudes, and personality traits In contrast with no books and only a few monographs (525,823,887) published in America on intelligence and vocational adjustment, and two text-books (385,94) and four significant monographs (588,589,201, 545) on aptitudes and vocational success, there have been two scholarly books (227,775), at least four significant monographs (279,145,189,793) and a number of important reviews of research published in the journals (77,111,798,800), all dealing with the nature and role of interests

Psychologists who have had other specialties have paid much more attention to other types of characteristics, Allport (898) and Thorndike (830) being among the few to study interests, through the Allport-Vernon Study of Values (see below) and various introspective techniques Clinical psychologists have tended to devote their energies more to the measurement of intelligence and to the diagnosis of mental defect and malfunctioning, students of individual differences have focused on abilities, and personologists have been challenged more by problems of the organization of personality (12,743,554) and by needs and drives (557) The genetic psychologists are perhaps an exception, as they have paid some attention to the development of play interests, in a type of study illustrated by those of Lehman and Witty (461,462,464)

It is worthy of note that, when these differing approaches to the psychology of individual differences have briefly met, the result has more often than not been confusion Thus Lehman and Witty loosed a broadside at vocational interest inventories, decrying their use in counseling on the grounds that interests are unreliable (463), but their evidence to that effect was based on *expressions* rather than on *inventories* of interests This is an important distinction which will shortly be made clear Because of the accumulation of unsynthesized material

on the nature and development of interests, these topics are dealt with at length in this section The role of interest in vocational adjustment will be considered later, in connection with the validity of specific instruments

Definitions There have been four major interpretations of the term interest, connected with as many different methods of obtaining data In an attempt to clarify thinking in this area the writer has (800) classified them as *expressions, manifestations, tests,* and *inventories* of interests Each of these is taken up in turn, to provide a framework for the subsequent discussion of the measurement of interest

Expressed interest is the verbal profession of interest in an object, activity, task, or occupation, Fryer (277) called it specific interest The client simply states that he likes, is indifferent to, or dislikes the activity in question There has been relatively little research in this area since Fryer's (277) detailed review in 1931, as shown in subsequent reviews by Carter (145) and by Berdie (79) The conclusion to be drawn from the later reviews is the same as that drawn by Fryer the expressed or "specific" interests of children and adolescents are unstable, and do not provide useful data for diagnosis or prognosis For adults however, the picture is somewhat more optimistic, for Strong (775 657) has shown that the constancy of responses to the 400 items in his inventory ranges from 52 6 percent for high school juniors after six years (reflecting the instability of expressions of interest just referred to) to 82 8 percent for women physicians after one day, showing that even specific or expressed interests are rather stable in adults over a short period The importance which may be attached to expressions of specific interests clearly varies with the maturity of the client As Gilger (290), Lurie (490) and Trow (875) have shown, it also depends upon the ways in which the questions are phrased, for some questions concerning vocational interest are so put as to elicit information concerning vocational *choice,* some to ascertain vocational *preferences,* and some to evoke vocational *fantasies* The degree of realism represented by the expression of interest varies with the type of question asked Studies of the relationship of expressed preferences to scores on Strong's Inventory, discussed below, illustrate this fact

Manifest interest is synonymous with participation in an activity or an occupation Objective manifestations of interest have been studied in order to avoid the subjectivity of expressions or to avoid the implication that interest is something static Thus Kitson (428 Ch 8) has urged that the verb "to be interested" should be used, indicating that a process

and activity are involved In this approach it is assumed that the high school youth who was active in the dramatic club has artistic or literary interests, and that the accountant who devotes two evenings per week to building and operating a model railroad system is interested in mechanics or engineering It is generally appreciated that such manifest interests are sometimes the result of interest in the concomitants or by-products of the activity rather than in the activity itself The high school actor may have merely been seeking association with others, which he may later need less or obtain by different means In other cases the opportunities for the manifestation of an interest may be limited by the environment or by financial considerations, so that an expressed interest has no manifest counterpart For these reasons, manifest interest has not been used as a predictor of interest in many studies, although it has often served as a criterion, the reasoning being that anything as dynamic as interest should in most cases find an outlet

Tested interest is here used to refer to interest as measured by objective tests, as differentiated from inventories which are based on subjective self-estimates It is assumed that, since interest in a vocation is likely to manifest itself in action, it should also result in an accumulation of relevant information Thus interest in science should cause a person to read about scientific developments, whether in a science course or in the daily paper, and to acquire and retain more information about science than would other people Fryer (277 664 ff) has reviewed the attempts which were made by O'Rourke, Toops, Burtt, McHale and others during and after World War I to measure interest by means of the amount and type of information retained, and has pointed out that these were not followed up because of the cumbersomeness of memory and information tests

With the improvement of testing and statistical techniques which subsequently took place, however, interest in the development of interest tests revived. At this time Greene published his Michigan Vocabulary Profile Test (308), measuring interest through specialized vocabularies The Co-operative Test Service brought out a general information test which Flanagan (262) described as a measure of interest in several areas The writer and his students at Clark University (806,805,574) began a series of investigations designed to develop an attention or recent-memory test of interest in vocational activities During World War II the Aviation Psychology Program brought together several psychologists who had been working along these lines (R N Hobbs, R R Blake, D E Super, J C

Flanagan, and F B Davis) Their efforts resulted in the development of a General Information Test which gave differential scores for pilots, navigators, and bombardiers and which proved to be the most valid single test in the Air Force's selection and classification battery (264,214) The writer constructed a similar test for the American Institute for Research which has been used in the selection of pilots for commercial airlines Other civilian applications are also being made, and the technique will in time probably prove to be generally useful for selection and counseling

Inventoried Interest is assessed by means of lists of activities and occupations which bear a superficial resemblance to some questionnaires for the study of expressed interests, for each item in the list is responded to with an expression of preference The essential and all important difference is that in the case of the inventory each possible response is given an experimentally determined weight, and the weights corresponding to the answers given by the person completing the inventory are added in order to yield a score which represents, not a single subjective estimate as in the case of expressed interests, but a pattern of interests which research has shown to be rather stable The apparently logical objection that no statistical combination of unstable elements can yield a stable total is met by Strong's study (775 871) of the effect of changes of responses to specific items on inventory scores although changes of expressions of liking or disliking of as many as 125 of his 400 items were found, these shifts had no appreciable effect on scores for occupational interests The reason for this is that shifts in one direction are balanced by shifts in the other direction, the underlying pattern or trend of interest being constant Strong's work provided a foundation for a great many studies in the psychology and measurement of interest, and made possible the development of practical instruments for use in counseling and selection He has summarized most of the significant research with his Vocational Interest Blank in a volume (775) which is one of the classics in the field of measurement Other inventories have been developed by Kuder (446), Garretson and Symonds (279), Dunlap (219,713), and others, some of these are discussed later in this chapter

The term interest is also used to convey other concepts, the most relevant of which are *degree of interest* or strength of motivation and *drive* or need The former needs no discussion, as it is a matter of degree rather than of kind when it is said that someone is vitally interested in attaining a goal, the statement is one concerning the degree of some

underlying (inventoried) interest or the strength of some drive. The concept of interest as drive does require discussion, for when it is said that an individual is interested in winning friends or in gaining prestige, the type of interest referred to is not covered by any of the concepts so far discussed Interests or drives of this type are of a different and more fundamental order than either specific or underlying interests, they constitute a deeper layer of personality Unlike interests, which are sometimes included under the heading personality and sometimes not, drives or needs are generally considered to be one of the central aspects of personality They are therefore discussed, together with their vocational significance and methods of measuring them, in the next chapter

Types of Interests As in the case of intelligence testing, progress in the measurement of interests was first made possible by a shotgun approach which was concerned less with the specific nature of that which was being measured than with the fact that it could be measured The all-important discovery made by Strong and his students was that the interests of men in a given occupation, e g , engineering, were different from those of men-in-general (775 Ch 7, 174) It was only after scales had been developed for the measurement of the interests of men in a number of occupations that factor analysis (836) and item analysis (446) revealed the nature of these interests For this reason the logical sequence of topics which follows is not the historical order in which discoveries were made

Interest factors were first studied by Thurstone (836), who applied factor analysis to 18 occupational scales of the Strong Vocational Interest Blank Strong (775 Ch 8 and 14) later made several factor analyses, in the last of which he used data from 36 occupational scales, first without rotating the axes (like Thurstone) and then by rotating them For clarity's sake, the results of these three analyses are presented in Table 27, together with data from three other studies and a logical synthesis of the findings of all six studies

Allport and Vernon developed their Study of Values as a measure of the values postulated by Spranger Lurie (489) also devised an instrument for appraising these values and, unlike Allport and Vernon, subjected it to factor analysis These two lists of factors are also presented in Table 27

Further evidence concerning the nature of interest factors is provided by Kuder's work with his Preference Record (described below) This inventory gives scores for nine types of interests,[1] which cannot be called

[1] A tenth, "Outdoor interest, has been added

TABLE 27

INTEREST FACTORS REVEALED BY SIX STUDIES AND LOGICAL SYNTHESIS

Thurstone	*Allport-Vernon*	*Lurie*	*Strong*		*Kuder*	*Synthesis*
			Unrotated	*Rotated*		
Science	Theoretical	Theoretical	Science	Science	Scientific	Scientific
People	Social	Social	People	People	Social-Service	Social-Welfare
Language			Language Things vs People	Language Things vs People	Literary (Mechanical)	Literary Material
Business	Economic / Political	Materialistic	Business	System	(Clerical) / Computational	System
				Contact	Persuasive	Contact
	Aesthetic				Artistic	Artistic
	Religious	Religious				
					Musical	Musical

factors in the statistical sense of the term as they were not isolated by factor analysis methods, but which amount to about the same thing as they are based on item analysis and are therefore internally consistent and mutually independent Kuder originally developed seven scales by this method, these are listed in Table 27 He later added two more, which are listed in parentheses because, unlike the others, they had substantial correlations with other keys mechanical interests correlated 405 with scientific, and clerical 50 with computational

Finally, after a study of the factors appearing in columns one through six of Table 27, together with the literature upon which they are based, the writer has developed the list of factors appearing in the last column of this table headed "Synthesis" The naming of statistically isolated factors is a highly subjective and arbitrary process For example, three authorities have variously named the same factor "interest in male association," "interest in order or systematic work," and "non-professional interests" (775 164–166) In one sense, therefore, the writer is justified in attempting a synthesis of the findings of various investigators and in applying his own names to the various categories, in another sense, the whole process of naming interest factors is open to criticism as a potentially misleading one It can be justified, perhaps, on the grounds that a cautiously named concept, cautiously used, is better than no concept at all it merely behooves the name-giver to point out the need for caution

Table 27 brings out complete agreement on the first interest factor, the *scientific*, which may be defined as an interest in knowing the why and how of things, particularly in the realm of natural science (only the

Allport-Vernon attempts to assess interest in *scientia* in the philosophical sense) There is agreement also on the second factor, interest in *social welfare* or in people for their own sake The third factor is not provided for by Allport and Vernon or by Lurie, who were limited by Spranger's postulates, but as factor analysis, like qualitative analysis in chemistry, can isolate only the elements which were originally put into the compound, the lack of positive findings in these studies can be disregarded The Thurstone-Strong-Kuder data can be accepted as evidence of the existence of a *literary* interest factor, consisting of interest in the use of words and in the manipulation of verbal concepts A fourth factor, again not revealed by the Spranger-inspired studies, by the Thurstone analysis (which was presumably based on too few occupations), nor by the Kuder procedure (though perhaps partly covered by his mechanical scale), but found in Strong's two analyses, might best be called the *material* or concrete, although Strong named it "things vs people" or, on the basis of negative loadings in the literary and linguistic occupations, "language" The writer prefers Kitson's term "material" because the occupations in which it has heavy positive loadings tend to involve working with tangibles Carpenter, mathematics-science teacher, farmer, printer, production manager, engineer, chemist (these last two have heavier scientific loadings), and even policemen and accountant may be included in this category, since they are concerned, respectively, with the protection and the management of property The fifth factor, one concerning which there is considerable agreement, is the *systematic* or perhaps record-keeping, it emerged most clearly in Strong's more refined analysis but he refused to name it, although he states that it might be called the C P A factor Kuder's computational interest appears to be similar, and it is probably covered also by Thurstone's business, Allport and Vernon's economics, and Lurie's misnamed philistine (or materialistic) values The sixth, or *contact* factor, is also probably included in the too comprehensive complex of factors called business and economic in the Thurstone and Spranger categorizations, refined by Strong's final analysis It is the second factor which Strong thought it wise to refrain from naming until more occupations were found to have loadings of it It seems to involve interest in meeting or dealing with people not for their own sakes but for material gain Kuder's persuasive interest appears to be identical with it Finally there are the *artistic* and *musical* factors, the former agreed upon by Allport and Vernon and by Kuder, and suggested in another study of Thurstone's (837), the latter isolated by Kuder

only and therefore quite tentative, although the failure of Strong and Thurstone to find such a factor proves little in view of the presence of only one musical occupation in their lists

Occupational differences in patterns of interests were, it has already been pointed out, the basic discovery which made possible subsequent studies of the nature and role of vocational interests Beginning his work in interest measurement as a member of the outstanding group of applied psychologists who were assembled at the Carnegie Institute of Technology after World War I, Strong continued to experiment with the vocational interest inventory technique after he joined the faculty at Stanford University, and there succeeded in establishing the fact that the inventoried interests of men who are engaged in different occupations differ significantly from those of men-in-general (775 Ch 7)

Some occupational groups, however, were not distinguishable from men-in-general, Strong's early attempts to develop scales for executives and for teachers failed (775 20,161 ff), and in his later studies of the interests of public administrators (779,780) he encountered difficulties which were essentially similar The reason for the failure to establish patterns of interest peculiar to executives and teachers, and for the lack of validity of the public administrator scale for some groups of administrators, may lie in the fact that these are not truly occupational groups Strong's work in the development of teachers' scales has shown, for example, that men social-studies teachers seem to be primarily social-welfare workers (r with YMCA secretary = 87), mathematics-science teachers resemble skilled tradesmen (r with carpenter = 68, with printer = 72), and the correlation between the interests of men in these two types of teaching occupations is practically zero (r = 13), as shown in Strong's table of intercorrelations (775 opposite 716) Similarly, the executive group was made up of men who were essentially engineers, lawyers, or other specialists (770), and the public administrators also included many men who were professional men at heart but who had been given administrative responsibility (779)

The occupations which were differentiable on the basis of interest patterns of the men engaged in them could be grouped, Strong found (775 Ch 8), according to the degree of similarity which existed between their interests Some of the occupational interest scales were positively intercorrelated, others negatively, in varying degrees Strong therefore grouped the various occupational scales on the basis of these intercorrelations, establishing 60 as the minimum intercorrelation necessary for two

occupations to be assigned to the same family The resulting families may be characterized as follows

Biological Science Occupations	e g Physician
Physical Science Occupations	e g Chemist
Technical Occupations	e g Printer
Social Welfare Occupations	e g Y Secretary
Business Detail Occupations	e g Accountant
Business Contact Occupations	e g Life Insurance Salesman
Linguistic Occupations	e g Lawyer

The terminology is essentially Darley's (189), but not that used by Strong, who has been extremely reluctant to name groups which seem at all heterogeneous He characterized the second group as "mathematics and physical sciences," the fourth as "handling people for their presumed good," the fifth as "office," the sixth as "sales," and the seventh and last as "linguistic" (775 160), but felt that the presence of such vocational groups as artists and architects in the first or biological science group (r artist-physician = 79) makes it difficult to name, and that aviators, carpenters, mathematics-science teachers, and policemen make odd bedfellows in the so-called technical group, even though their intercorrelations with the printer scale are 65, 73, and 72 respectively As Strong points out in his discussion (775 159–160), the sub-professional technical group appeared originally as part of a general scientific group which included also the biological and physical science occupations, but which broke up into the three scientific or near-scientific groups of the current classification when more occupational scales were devised As additional occupational scales are developed, it is probable that the so-called technical group will further subdivide, an hypothesis for which Strong provides some important substantiating data in the analysis of the effect of the point of reference (775 Ch 21 and 22)

Another group which seems likely to subdivide as more keys are added is the contact or sales group The danger involved in either of these names is brought out by the fact that public utility salesmen belong more in the business detail group (r office worker = 69) than in the contact family (r life insurance salesman = 39) Strong therefore feels that there will in time be a new sales group, consisting of house-to-house salesmen The classification of occupations on the basis of interests must

therefore be considered tentative, and one must not let the very natural desire to give names to categories lead to the making of false generalizations

The data on differences between kinds of teachers and salesmen raise a question concerning other occupations They prompt one to ask whether a sufficiently refined analysis would reveal similar differences between mechanical, electrical, civil, and chemical engineers, for example, or between various types of secretaries in the YMCA Using his standard techniques of comparing one occupational group with men-in general (described in the section dealing with the inventory), Strong found no differences between the interests of the various types of engineers, the correlation between civil and electrical engineer, for example, being 86 (775 118) He obtained similar results with scales for the interests of YMCA general and boys' work secretaries, only the physical directors being distinct enough (r combined Y secretary scale = 74) to warrant a separate key These results would seem to point to the conclusion that some occupations can be broken down into specialized subgroups on the basis of interests, and that others cannot To the first category one might add teachers and public administrators, already discussed in another connection, and certain types of sales work, to the latter, sales managers and salesmen in certain fields such as life insurance and vacuum-cleaners It would be interesting to know the facts for criminal and corporation lawyers, surgeons, pediatricians, and psychiatrists [surgeons do not differ from physicians (775 697)], and clinical, industrial, differential and physiological psychologists A study carried out under Paterson's supervision at the University of Minnesota deals with this last occupation In view of the evidence accumulated by Strong it seems safe, in the meantime, to state that when compared with the interests of men-in-general, the interests of men in a broadly defined occupation are so similar as to obscure differences between specialties A mechanical engineer, when compared to non-engineers, is more like unto than different from a civil or electrical engineer, for then the common factor, engineering, is crucial

It has been shown, however, that when a different point of reference is used, men in specialties within an occupation can be differentiated Using engineering students as subjects, Estes and Horn (241) compared the interests of each type of engineer, mechanical for example, with those of all other types of engineers studied The point of reference in this

study was therefore not men-in-general, but engineers-in-general Under these conditions the differences between the interests of the various specialty groups became visible, and separate scales could be developed Strong recognized the possibilities of this approach (775 120), but has not attempted to capitalize them, nor has anyone else The method is one which might well commend itself to professional schools interested in providing better guidance services for their students or in improving their selection procedures

The point of reference used in constructing occupational scoring keys for interest inventories has been found to have one other important effect on our knowledge of occupational differences in interests The first men-in-general group used by Strong consisted, for reasons of convenience, of men for whom test data were on hand and who were not in the occupation under investigation (775 555) This happened to be an economically somewhat select group, for the first scales constructed were for occupations which were of either professional or managerial calibre When Strong's Vocational Interest Blank was used with men from the lower half of the occupational heirarchy, little differentiation of interests was found printers, carpenters, policemen, and farmers, coming from the skilled trades level, had so much in common, when compared with a professional-managerial-clerical men-in-general group, that the differences between them were not very significant (the intercorrelations approximate 70), and persons habitually employed at the semiskilled levels seemed to be undifferentiated on the basis of their interests (84)

These findings led the staff of the Minnesota Employment Stabilization Research Institute to hypothesize concerning the differentiation of semiskilled and unskilled workers on the basis of interests (84), and prompted Strong to pursue a line of research already suggested by his work with the women's blank (775 554) He therefore developed occupational scales based on three different points of reference These consisted of (1) business and professional men earning $2500 or more per annum (rather like the original scales), (2) a proportional sample of all occupational levels averaging, like the general population, at the skilled trade level, and (3) a proportional sample of skilled, semiskilled, and unskilled workers, averaging at the semiskilled level For convenience these three reference points, called P_1, P_2, and P_3 by Strong, may be referred to as the white-collar, general, and blue-denim groups Three hypotheses were set up to be tested by means of these occupationally similar, but referentially different, keys These were

1 Certain occupations at different levels have the same types of interests (e g , engineers at the professional and mechanics at the skilled),
2 The rank and file cannot be differentiated by their interests (e g , semiskilled workers tend to make no high scores),
3 Men in the lower-level occupations have their own occupationally-specialized interests (e g , when compared to other semiskilled workers, drill-press operators have interests which are different from those of electrical-unit assembly workers)

Using scales based on the general point of reference (P_4), Strong found that the correlation between the printer and carpenter scales was .24, whereas it was .73 when the white-collar point of reference (P_1) was used, similarly, the correlation between printer and policeman was −.27 instead of .59 In other words, when an appropriate point of reference is used, differences between the interests of men in a given occupation and those of men in the reference group appear significant, when an inappropriate reference point is used, differences between the interests of men in the occupation being studied and those of "men-in-general" are obscured This holds whether it is men in a low-level occupation being studied against a high-level men-in general group, or men in a high-level occupation being studied with a low level men-in-general group as point of reference Strong's third hypothesis was therefore confirmed, and it is to be expected that in due course occupational interest scales will be developed which will be useful with men of less than average socio-economic level and for counseling and selection for more occupations at the skilled and semiskilled levels, many of which should be found to have differentiating patterns of interests

Lest it appear from the preceding paragraphs that all of our knowledge concerning the differentiation of occupational groups on the basis of interests is based on work with Strong's inventory, it should perhaps be mentioned that Kuder (446) and Triggs (in an unpublished paper) have confirmed Strong's general findings for some forty occupations with Kuder's Preference Record Triggs went so far as to establish differential interest patterns for various types of nurses, including supervisors and public-health nurses The Allport-Vernon Study of Values has shown similar trends with pre-occupational groups in colleges (216), but has not been much used with men and women actually engaged in occupations Reference has been largely to work with Strong's Blank simply because,

as an older and more thoroughly studied vocational interest inventory, it provides more data from which to draw conclusions

Socio-Economic Differences The preceding discussion of the effect of the point of reference on the identifiability of patterns of interests was virtually a treatment of research which has been carried out on socio-economic differences in occupational interests, at least in so far as methodology is concerned There still remains the task, however, of describing the differences in interests which characterize the various occupational levels The relevant work is reported in Strong's book (775 Ch 10), in connection with his scale for measuring occupational interest level, or the socio-economic level at which an individual would be placed on the basis of similarity of interests Men who are successfully employed in the higher level occupations tend to have more interest in literary and legal activities and in business contact work, and less social welfare and sub-professional technical interest, than men in lower level occupations Men in legal and literary occupations, salesmen, and scientists tend to make high occupational level scores, although there is no relationship between the scientific and occupational level scales on Strong's Blank Senior public administrators score higher on occupational level than do junior public administrators (779) Strong suggests that the scale measures managerial ability

On the other hand, it has been more plausibly suggested by Darley (189 60 and 66) that occupational interest level is indicative of aspiration level, that it "represents the degree to which the individual's total background has prepared him to seek the prestige and discharge the social responsibilities growing out of high income, professional status, and recognition or leadership in the community, at the lower end of the scale, the individual's background has prepared him for the anonymity, the mundane round of activities and the followership status of a great majority of the population" He suggests, also, that those who are characterized by a low level of occupational interest are likely to lack the motivation which results in staying power in college Kendall (422) has attempted to validate this hypothesis with three groups of 100 men each at Syracuse University, selected from the entering freshman class on the basis of high, average, and low occupational level scores on Strong's Blank These three differing occupational level groups were found to differ also in mental ability as measured by the Ohio State Psychological Examination Those who were high on these measures made higher hour point ratios during the first semester When intelligence was held

constant the academic achievement of the three occupational level groups was again found to differ, the differences being significant at between the one percent and the 5 percent levels The differences are therefore not completely clear cut, but they do suggest that those with extremely low occupational interest levels are likely to find college work foreign to their taste, whereas it will be congenial to those who are characterized by high occupational interest levels

Avocational Differences What has been found true of occupations has been found to apply also to avocations In a study of model engineers, amateur photographers, amateur musicians, and stamp collectors, the writer (791) found that men who were active in the first three avocations had patterns of interests which differentiated them from each other, and that the first two interest patterns resembled each other ($r = .58$) whereas the first and third had nothing in common ($r = .02$) Although the number of avocations studied is small, this suggests that they are differentiated and may be classified in ways similar to occupations The interest patterns of stamp collectors were found to be similar to those of other groups of men, suggesting that philately is an avocation which, like the vocation of executive, cuts across basic interest patterns which are much more important than the interest common to men engaging in it It is noteworthy, also, that the three differentiable avocational interest patterns resemble those of the expected occupations, e g , the model engineers have interests like those of professional engineers, whereas the interests of stamp collectors are difficult to classify vocationally

Sex Differences Popular stereotypes as to the masculinity and femininity of interests are widespread, and it is natural to ask what research in the psychology of interests found in this area Studies made by Terman and Miles (820), Carter and Strong (148), Yum (952), Strong (775 Ch 11), Kuder (446 23), and Traxler and McCall (868) All agree that men tend to be more interested in physical activity, mechanical and scientific matters, politics, and selling Interest in art, music, literature, people, clerical work, teaching, and social work is more characteristic of women It is especially worthy of note that masculinity and feminity are scaled traits rather than dichotomies people are not masculine or feminine in their interests, but more or less masculine or feminine Some men are very masculine, and so are some, but fewer, women, some women are very feminine, and so are some, but fewer, men It is interesting to speculate as to whether the higher incidence of cultural (artistic, literary, musical, and social) interests in women means that they are constitutionally the

carriers of culture, or whether they have simply taken on that role because nature forced men, as the stouter animals, to take on the competitive, constructional, and provisioning roles Anthropological studies suggest the latter, since there are a few societies in which men are the domestics and women the providers But physical constitution seems to play a part, as shown by the preponderance of active-male societies A good illustration is Miles' (529) case study of a boy raised for 17 years as a girl despite the seemingly overwhelming feminine influences to which he was subjected, he made definitely masculine scores on the Terman-Miles Masculinity-Femininity Test and on Strong's Vocational Interest Blank (scored for masculinity-femininity of interests)

Age Differences Counselors and psychologists who have not carefully studied the literature on change of interest with age frequently question the wisdom of giving much weight to measures of interests because of the possibility of change of interest with age This question overlaps to some extent with that of the permanence of interests, discussed below, but it is distinct in that it focuses on the relationship between age and change, rather than on the effects of experience

Three important studies have been made of the differences in interests which are associated with differences in age The first of these was by Strong (771), incorporated and brought up to date in his later book (775 Ch 12–13), the second was a series of follow-up studies by Strong (775 358–362), the third was part of the Adolescent Growth Study of the University of California, written up in a series of articles by Carter and others and summarized in his monograph (145) and in a journal article (144)

Strong's first approach consisted of comparing the interests of men at ages 15, 25, 35, and 55, both by analysis of individual items (ages 15, 25, and 55) and by the construction of interest-maturity scales for each of the four age levels selected for study These analyses revealed that age differences are less significant than occupational differences The interests of 15-year-olds agree in large measure with those of 25-year-olds ($r = .57$), are more like those of 35-year-olds ($r = .66$), and even more like those of 55-year-olds ($r = .89$) (775 279), as about one-third of the change that takes place between ages 15 and 25 occurs during the first year (15 5 to 16 5), one-third during the next two years, and one third during the next seven years (775 259) it is clear that interests are fairly well crystallized by age 18 Boy's interests tend to become less like those of physicians, dentists, and engineers as they approach age 25, and more like

those of office workers, salesmen, accountants, physical directors, social science teachers, and personnel managers, those whose interest-maturity scores on Strong's Blank are high are least likely to show changes of interest patterns, whereas the interests of those whose interest-maturity scores are low are most likely to undergo change

The slight changes that take place after age 25 tend to be an undoing of those that took place prior to that time, as shown by the higher correlations between the interests of 15 and 35 or 55-year-olds than those of 15 and 25-year-olds already cited Strong has confirmed this with two different sets of data (775 283–285) A study by Sollenberger (726) provides a basis for the conclusion that increases in hormone activity in adolescence account for the changes that take place in boys at that stage, perhaps it is decreases in hormone activity after the mid-twenties [suggested by studies of sex habits conducted by Kinsey (424)] which account for the reversal This tendency toward an undoing of the 15 to 25-year-old changes should not, however, be interpreted as a reversal of all trends, for the decreased interest in physical activity and daring continues beyond age 25 and is the most striking change during that period of little change, others are a decreased interest in occupations involving writing, and a lessened liking for change or interference with established habits Strong summarizes his work as follows "The primary conclusion regarding interests of men between 25 and 55 years of age is that they change very little When these slight differences over thirty years are contrasted with the differences to be found among occupational groups, or between men and women, or between unskilled and professional men, it must be realized that age, and the experience that goes with age, change an adult man's interests very little At 25 years of age he is largely what he is going to be and even at 20 years of age he has acquired pretty much the interests he will have throughout life" (775 313)

The second series of studies conducted by Strong were follow-ups of 175 Stanford freshmen, retested nine years later, and of 168 Stanford seniors, retested ten years after graduation The average correlation between test and retest scores was 56 for those first tested as freshmen (ages 18 and 27) and 71 for those first tested as seniors (ages 21 and 31) These findings from longitudinal studies confirm the conclusions drawn from Strong's cross sectional analyses in revealing a fair degree of permanence of interests in 18-year-olds, and a substantial degree in 21-year-olds The lowest retest reliabilities at these ages were in the social welfare occupations, and the highest in the scientific and literary occupations,

that this is partly due to decided increases in social welfare scores and relative stability in literary scores is shown by critical ratios ranging from 2 5 to 5 1 for the test retest means of the former, and by critical ratios of −0 6 and −1 5 for those of the latter (775 363) As those for the test-retest means of the scientific occupations ranged from 0 6 to 4 2, showing a tendency for some increase in scores to take place there, it must be deduced that the changes in scientific scores are regular and do not generally affect the rank order of the persons tested, while the changes taking place in social welfare interests are irregular and do generally affect the rank order of the persons tested In other words, those who make the highest scientific scores tend to remain highest and those who make the lowest scientific scores as seniors tend to remain lowest, while some of those who made the lowest social-welfare scores make substantial gains in this area and others do not Strong s other findings show that it is those persons with the lowest interest maturity scores who make the most radical gains

The Adolescent Growth Study investigations were, as the title implies, longitudinal studies of high school pupils who were tested with Strong's Blank in the 10th or 11th grades and retested each year until graduation from high school at about age 18 and, in some cases until after graduation from college These studies showed that the correlations between interest patterns in 10th or 11th grade and the last year of college are about as high as those between interests in the first year of college and five years after graduation, for Taylor's study (814) revealed a mean correlation of 52 for 11th graders retested six years later, as compared with Strong's average correlation of 56 for college freshmen retested nine years later Carter (144) and Taylor and Carter (814) have similarly demonstrated that the interest patterns of high school boys and girls (in practically the only studies of change in girls) remain fairly stable throughout the high school years Carter concluded that "the Strong scales are almost, but not quite, as reliable and stable when used at the high school level as when used with adults" and Taylor stated that "vocational interests, as measured by the Strong inventories, appear to be almost as permanent during the high school years as during adult life" It may be well at this point, however, to remember Strong's more cautious conclusions, already quoted on page 391

Communality of Interests The significance of differences in the interests of occupational, socio-economic, avocational, sex, and age groups obscures an important fact brought out by Strong's research (775 Ch 6),

namely, the fact that people's interests are far more similar than different, regardless of sex, age, or occupational status It is not really surprising to learn that people are human, and yet the fact is easily lost sight of when they are studied as men and women, boys and men, or professional men and skilled workers The likes of college men and women are very similar ($r = 74$), those of 15-year-old-boys and 55-year-old men are no less similar ($r = 73$), and those of unskilled workers and professional-managerial men resemble each other even more closely ($r = 84$) Underneath the very real differences among various groups of people we find an even larger common core which is of great social and philosophical importance

Stability of Interests The question of the permanence of interests is closely tied up with that of change of interests associated with age We have seen that age changes do take place in adolescence, but that the patterns of interests which begin to manifest themselves by age 15 tend to be those which are revealed at ages 25, 35, and 55 Most of the change which does take place with maturity is complete by age 18, the type of change which may take place at that age is systematic and predictable on the basis of interest inventory data (interest-maturity scores) It is still pertinent, however, to inquire concerning the permanence of interests when they are subjected to influences which may change them in one direction or another Kitson (430), for example, has described a series of projects designed by O'Rourke to modify interest in vocational activities The evaluation in terms of changes in *expressed* interests showed that pleasant experiences do change overt attitudes toward activities But whether or not underlying interests, or interest patterns as measured by Strong's Blank, are thereby modified remains to be ascertained

There has been surprisingly little study of this problem, insofar as inventoried interests are concerned, the focus has generally been on the effect of rather limited experiences on expressed preferences However, Burnham (125), Glass (775 379), Mather (775 379), Ryan and Johnson (660), Klugman (435), Strong (775 388–411), and Van Dusen (889) have investigated the effects of school and vocational experiences on inventoried interests

The relationship of change of inventoried interests to college grades among Yale students was studied by Burnham, who found no relationship, such changes in interests as did take place could not demonstrably be attributed to the kind of grade achieved in college courses Klugman's

contrary findings concerning the clerical interests of high school girls probably prove little, in view of the general tendency for girls and women to make high clerical scores Van Dusen worked with engineering students at the University of Florida, a group whose mean scores, as Strong points out (775 278), were very low, suggesting that they may have been, not a selected group, but rather a heterogeneous collection of state university freshmen who thought they would like to study engineering. He found a slight and statistically insignificant decrease in the retest scores of students who had given up their freshman choice by their senior year, and similar increases in the engineering scores of students who remained in that field throughout college Strong failed to find the last trend in Stanford engineering students, but confirmed the others in studying the occupational histories and test scores of his Stanford seniors who were followed up ten years later those who were finally employed in a field other than that preferred when they were seniors made retest scores which were 3 6 standard score points lower than their original scores in the latter field, the critical ratio approached significance (2 5) The retest scores on the finally-entered occupation were higher by a comparable amount than were the original scores for that occupation It is significant that there were no changes in the scores of those who entered and remained in the field of their preference as seniors ten years of occupational experience did not increase interest in the field of employment Strong also analyzed the employment histories and test scores of Stanford freshmen retested nine years later, and found essentially the same results

Mather, as reported in Strong, found no increase in home economics teacher scores after practice teaching in that field (a limited sample of experience, and a group already somewhat selected by training), she did, however find substantial increases (4 7 standard score points, or one-half sigma) in the appropriate interests of 45 students who were retested after their first two years of exposure to the field of home economics These studies suggest, either that experience in a field *inappropriate* to one's interests causes one to become even less interested in that field (and, conversely, more interested in appropriate occupations), or that it helps to bring about a better understanding of one's likes and dislikes and obtain more nearly true scores on an interest inventory In the case of *appropriate* experience, however, there seems to be no effect, perhaps because understanding is already good enough not to be affected As

the inventory is a self-portrait technique, the second explanation seems acceptable

Not in keeping with this interpretation are Glass' results, which showed that the interests of unselected engineering freshmen who remained in engineering college until graduation became less like those of engineers (shift from B+ average to B) while the interests of those who dropped out as freshmen were interpreted as having become somewhat more like those of engineers. In the latter instance, however, it was an insignificant raw score increase of two points, but one which happened to change the mean letter grade from B to B+ The decline in the interest of the graduates may have been due to poor guidance and selection, such as frequently results in many able but uninterested students persisting until graduation thus many graduate engineers never enter engineering occupations, but become salesmen, accountants, etc

Two studies have considered the relationship between length of time in an occupation and similarity of interests to those of men successful in that field Both of these (660,775 487) found insignificant correlations (00 to − 12) between the interest scores and length of experience of sales and service men in the one case, and of life insurance salesmen in the other

It is perhaps not so difficult to synthesize these findings into a theory of the effect of adolescent and adult experiences on vocational interests as their occasional apparent discrepancies suggest Strong (775 380) concludes that "the interests of occupational groups are present to a large degree prior to entrance into the occupation and so are presumably a factor in the selection of the occupation," rather, the implication being, than the result of experience in that occupation This conclusion is legitimate and adequate enough as a generalization concerning the permanence of interests, but it does not go as far as the data warrant in describing the modification, as opposed to the creation or destruction, of interests by experience. Before citing the attempts of others to provide such a synthesis and interpretation, however, three more aspects of the problem of the origin and development of vocational interests need to be dealt with These are family resemblances, and the roles of aptitudes and of personality factors.

Family Resemblances The inventoried vocational interests of 110 pairs of fathers and sons were correlated by Strong (775·680), the sons ranging in age from 15 to 28 with a mean of 22 years The range of

correlations for 22 vocational interest scales was from 11 to 48, the average intercorrelation being 29, the average intercorrelation for randomly assorted men and boys, from the same total group, was 03 The interests of 125 pairs of fathers and sons were studied by Forster in a thesis cited by Berdie (79 145), in this study the sons were all students at the University of Minnesota The range of intercorrelations for 25 occupational interest scales was from 00 to 48, with an average of 33 Berdie (77) found that the sons of men in the skilled trades and in business tended to have inventoried interests in those fields, although the relationship did not hold for other fields The reason may lie in the fact that, as shown in a study of the writer's (790), these two occupational fields are near the top of the blue-denim and white-collar occupational ladders, making it socially acceptable for sons of business men and skilled workers to aspire to emulate and identify with their fathers, but less easy for the sons of unskilled, semiskilled, or clerical workers, who are not at the top of either ladder, to do so It would be difficult to explain the lack of relationship among the interests of professional men and their sons in Berdie's study in terms of this hypothesis, since they are already high on the white-collar ladder, were it not for positive results which he reports from a study by Dvorak She found that the interests of physicians and their sons were similar This suggests that sampling errors may have affected Berdie's results for this one occupational level, in which case the hypothesis that family resemblances are most likely to be found at the levels which are considered near the top of a social ladder would be confirmed

Other family relationships studied are those of twins, both identical and fraternal, in a report by Carter (142) His subjects were 120 pairs of twins, 43 of the pairs being monozygotic For these latter the average correlation was 50, whereas that for dizygotic twins was 28 Carter, Strong, and others have argued that the closer resemblance of the interests of identical than of fraternal twins does not prove that heredity plays a part, for "the environments of identical twins are more similar than those of fraternal twins" (145 51) This is an oft repeated statement, but one which has not, to this writer's knowledge, ever been demonstrated It is even more logical to maintain that the environments of fraternal twins are more similar than those of fathers and sons, in view of the differences in age, generation, and daily routines in the latter case, *but we have seen that the interests of fathers and sons resemble each other just as closely as do those of fraternal twins* (r = 29 or 33 and 28 respec-

tively) It seems necessary, then, tentatively to conclude that the greater similarity of the interests of identical twins, as contrasted with those of fraternal twins, is not due to the *potentially* greater similarity of their environments, but rather to the *demonstrably* greater similarity of their heredities

This viewpoint is that espoused by Strong (775 682), who points out that if environment is so predominantly important it is odd that boys and girls learn different interests by the time they are 15 years of age and unlearn them so little thereafter (see the discussion of sex differences), and that occupation-like differences which are found in the interests of adolescents are affected so little by subsequent training and experience.

Aptitude as a Source of Interest The necessary conclusion, as Strong sees it, is that "interests reflect inborn abilities" (775 882) There is little evidence, however, by means of which this inductive hypothesis can be verified or rejected It has been demonstrated that there is some relationship between intelligence and inventoried interests Strong (775 332–333) has summarized the various studies, showing that the correlations range from about − 40 to 40 depending upon the type of interest The positive correlations are with scientific and linguistic interests, while the negative relationships are with social welfare, business contact, and business detail interests Readers who happen to be social workers or teachers need take no offense at the first of these negative relationships, which shows that no normal person is too dull to take an interest in his fellowmen, and that there is a tendency for mentally superior people to let themselves become absorbed, perhaps to too great an extent, in other matters As scientific and linguistic occupations deal primarily with abstractions, and social welfare and business occupations at least partly with tangibles, what these relationships demonstrate is that, without the ability to understand, there can be little genuine interest

There have been fewer studies of the relationship between special aptitudes and inventoried interests. Adkins and Kuder (8) correlated scores on the Primary Mental Abilities Tests with those on the Kuder Preference Record, and found that only one correlation was above 30 that between number ability and computational interest in women Although this one relationship seems logical, it did not hold for men, and other equally appropriate relationships were not found in a high enough degree to justify any positive conclusions concerning the relationship of aptitude to interest However, Darley (191) found somewhat clearer indications of relationships between PMA Test scores and six

representative Strong scales (r's ranged from — 04 to .31), and Long (478) found the expected relationships with the Stanford Scientific Aptitude Test Other comparable data were reported in a thesis by Leffel (460), who found positive relationships (r = .46 and .42) between the O'Rourke Mechanical Aptitude Test and engineering and chemical interests on Strong's Vocational Interest Blank, and negative relationships between O'Rourke and Strong's social science teacher and lawyer scales (— 25 and — 25) and by Holcomb and Laslett (375), who obtained similar findings with Stenquist's mechanical (paper-and-pencil) test As the O'Rourke is to an indeterminate degree a measure of information, and therefore of interest as well as of aptitude, it would be difficult to draw any pertinent conclusions from Leffel's findings were it not that Holcomb and Laslett (375) and Moore (536) cite comparable data for the MacQuarrie and the Bennett Mechanical Comprehension Test It seems, then, that there is some relationship between aptitudes and interests

As so little research has been carried out to test Strong's hypothesis concerning the relationship of aptitudes to interests, it may be well to reproduce his reasoning on this point "An interest is an expression of one's reaction to his environment The reaction of liking-disliking is a resultant of satisfactory or unsatisfactory dealing with the object Different people react differently to the same object The different reactions, we suspect, arise because the individuals are different to start with We suspect that people who have the kind of brain that handles mathematics easily will like such activities and vice versa In other words, interests are related to abilities and abilities, it is easy to see, can be inherited There is, however, a pathetic lack of data to substantiate all this" (775 682–683) Strong believes that there are two reasons for this interests must reflect the environment, and they are evaluated by the environment Whereas a primitive Indian boy with fine finger dexterity might make arrowheads, the concomitant satisfaction of finger dexterity might make an urban American boy aspire to the occupation of dentist or watch repairman Interpreting this aspiration in terms of socio-economic levels, the professional man's son might want to be a dentist, the son of a skilled tradesman might want to be a watchmaker Establishing a causal relationship between aptitude and interest is difficult under these circumstances.

A conclusion diametrically opposed to Strong's was reached by Berdie after reviewing a number of studies of the relationship between ability and vocational interests, most of which actually dealt with choices or

expressed preferences rather than with inventoried interests (79 142). He wrote "The available evidence indicates, however, that a person's ability is not a very important factor in determining his interests, and although a relationship can be found between the two factors, this relationship is so small that we must look further if we are to understand the sources of vocational interests"

With so little evidence on which to base conclusions, it seems likely that this disagreement is one of orientation rather than of interpretation· each writer sees essentially the same facts, but, as the situation is not clearly structured, they are differently interpreted Berdie's orientation seems to be similar to that of Darley (189· Ch 6), who rejected Strong's deductions because "in general the magnitude of such correlations is too low to substantiate the hypothesis," and because differing amounts and kinds of aptitudes "might be required" for success in architecture and in chemistry, both of which belong to the same interest family. Concerning Darley's first objection, we have just seen that there is insufficient evidence, and that it assigns a real role to intelligence Concerning his second objection, one can only point out that it is hypothetical and that it would be quite logical for two partly overlapping complexes of aptitudes to contain some differing factors, thus resulting in two related but not identical constellations of interests such as architecture and chemistry, which belong in the same occupational interest group and which both require number and spatial aptitudes The different channeling in architecture and chemistry could be due to other aptitudes, social approval, or personality factors Darley's objections to Strong's hypothesis therefore do not seem very compelling Be this as it may, he thought it necessary to change the point of departure After also rejecting a form of recapitulation theory, because of general lack of substantiation of such theories, he writes "The adjectives by which our behavior is characterized in the description of others have usually been applied as personality values or attributes in late adolescence or young adulthood Our occupational stereotypes of the 'typical salesman' or the 'meek bookkeeper' or the 'absent-minded professor' evoke a series of such adjectives when we attempt to define the stereotype It is possible then that occupational selection and elimination is based on personality type as well as amounts and kinds of ability and aptitude. The third hypothesis of the origin of occupational interest types is that they are by-products of the development of personality types" (189 56) Darley goes on to cite evidence which he believes substantiates his hypothesis, quoting Carter

(144) to the same effect Such evidence is reviewed in the paragraphs which follow

Personality and Interests Social attitudes are the least fixed of personality traits in the sense of being most clearly and readily affected by the environment In a preliminary study of the relationship between these and vocational interests during the Depression, Darley (188) found that students with interests like those of personnel managers and YMCA secretaries had the highest morale, and those with interests like those of engineers and chemists were lowest on morale, as these are respectively measured by Strong's Blank and the Minnesota Scale for the Survey of Opinions. Such results in a preliminary study led to an analysis of data from 1000 cases tested at the University of Minnesota (189 63–65) This revealed that, contrary to the findings of the preliminary study, there was no relationship between morale scores and type of interests On the other hand, differences in liberalism and social adjustment were found those with welfare interests were most liberal, those with business interests least so, those with social welfare and business contact interests were best adjusted socially, those with linguistic and technical interests least so.

If values are thought of as representing a layer of personality which is deeper than those at which vocational interests and social attitudes are found, then it is significant that there is also some relationship between these two types of interests Sarbin and Berdie (668) obtained Strong and Allport-Vernon scores from 52 college students, and found positive relationships between scientific interests and theoretical values, welfare interests and religious values. Duffy and Crissy (217) obtained similar data from 108 college women, reporting intercorrelations which were in the expected directions and generally in the 30's Burgemeister (124) confirmed these findings with another group of 164 college women, reporting that the interests of librarians, artists, and authors, for example, tend to be associated with aesthetic values, and that those of physicians and science teachers tend to go with theoretical values Ferguson, Humphreys, and F. W Strong (253) have also confirmed these trends, with 93 college men

Personality traits at a somewhat deeper level were also included in Darley's investigation (189 63) These were measured by means of the Bell Adjustment Inventory and the Minnesota Scale for the Survey of Opinions, the former yielding scores for home and emotional adjustment and the latter for feelings of inferiority and family adjustment which

are of interest to us here Darley reports that home and emotional adjustment were not related to any occupational interest patterns; inferiority feelings were somewhat less common in those with welfare interests than those with a technical or no primary interest patterns, and family attitudes were somewhat better in men with business detail interests than in those with linguistic or no primary interest patterns, but neither inferiority feeling nor family attitudes differentiated between other interest groups Berdie (77) used the Minnesota Personality Scale and Strong's Inventory, and found that high school seniors with interests like those of engineers had inferior social adjustment, whereas those with social welfare interests were better adjusted socially and emotionally. In the only other study of this type known to the writer, Alteneder (14) found no correlations which exceeded 25 between men's adjustment and interest scores for six occupations, and only four which exceeded that for seven women's occupations These latter were 39 and 38 between social adjustment (Bell) on the one hand and linguistic and social work interests (Strong) on the other, and 34 and 26 between emotional and home adjustment on the one hand and teaching interest on the other Although the results of Alteneder's women's study are intriguing, the lack of positive results for the men's occupations makes them merely suggestive

A still deeper level of personality organization was studied by Triggs (873), who correlated cyclical, paranoid, schizoid, and other temperament traits as measured by the Minnesota Multiphasic Personality Inventory with vocational interests measured by the Kuder Preference Record Significant relationships reported in that paper were, for 35 college men, those between depression and social service ($r = -34$ and clerical (36) interests, psychopathic deviation and mechanical interests (-41), femininity and mechanical interests (-37), paranoia and computational (-42) and scientific (-38) interests, psychasthenia and scientific (-33), musical (33), and clerical (33) interests, and schizoid trends and musical (39) and clerical (32) interests It is perhaps worth noting that these relationships are suggestive of more positive personality adjustments being found with mechanical, computational, scientific, and social service interests, and of more maladjustments being associated with musical (psychasthenic and schizoid) and clerical (depressed, psychasthenic, and schizoid) interests These relationships are all significant at the 5, and occasionally almost at the 1, percent levels When the same techniques were applied to women college students, 60 in number, no relationships

were found except between lie score and musical and social service interests (the mean lie score for the whole group was normal) The apparent discrepancy between the sets of data for men and for women may be the result of the small size of the samples, which certainly require confirmation with larger numbers, but it is also possible that certain vocational interests could have pathological significance in men and yet be quite wholesome in women Triggs, at least, felt that the relationships which she found were significant

As Darley has pointed out, Terman and Miles', and Strong's data on masculinity and femininity of interests indicate a relationship between temperament and vocational interests, the endocrine basis of which has been demonstrated by Sollenberger (726) Work with an information test designed to measure temperament factors through information and interest (316 Ch 14,25, 925 68–74) tends further to substantiate the hypothesis that interests are related to temperamental factors.

Origin and Development of Interests The first published attempt to synthesize findings such as those reviewed above into a theory of the development of vocational interests was made by Carter (144), with a focus which is primarily environmental As he sees it, the individual derives satisfaction from identification with some group, by which means he attains status If his abilities permit, this identification is strengthened, if insurmountable obstacles are encountered, the process of identification is interfered with, the self-concept is changed, identification with another group must take place, and with it a new pattern of interests is developed which is more compatible with the aptitudes of the person in question Carter goes on to state that the interest patterns of adolescents tend to become increasingly practical, that in the beginning many adolescent interest patterns provide very unsatisfactory solutions of the problem of adjusting their aspirations to personal abilities and social demands He writes (144 186) "In this process of trying to adjust to a complex culture, the individual finds experiences which offer some basis for the integration of personality The pattern of vocational interests which gradually forms becomes closely identified with the self The pattern of interests is in the nature of a set of values which can find expression in one family of occupations but not in others"

This is essentially the line of thought developed independently by Darley, who quotes Carter in a briefer discussion of the same subject (189 57), and subscribed to by Berdie (79). This writer sees three serious defects in it

First, it is based partly on an environmentalistic interpretation of Strong's, Carter's, and Berdie's data on family resemblances, interpretation which, as we have seen above, does not seem warranted when viewed objectively, however laudable it may be to believe in the essential modifiability and improvability of man

Secondly, although it takes into account the role of aptitudes, personality is postulated as the basic factor, modified by the interaction of aptitudes and environment But this, we have seen, is on the basis of evidence which is fragmentary, tentative, and not much more convincing than that on the role of aptitudes which caused Strong to postulate that aptitudes are the fundamental factor

Thirdly, although Carter's description of the process of identification, trial, disruption, and reshaping of the identification sounds convincing, there is no evidence for it in the intensive analyses which he and other members of the California Adolescent Growth Study have published, nor in the publications of Darley, Berdie, and other Minnesota psychologists On the contrary, we have seen that everything that has been published on the development or stability of interests from the beginnings of adolescence on suggests that the form in which interest patterns begin to crystallize is essentially the form in which they remain, except as they are modified by glandular changes associated with age

An explanation of these phenomena which attempts to take into account the stability of inventoried interests has been advanced by Bordin (111). As he puts it, "One of the major facts which Strong has established concerning his blank is the continuity of interest patterns In general he has found that these patterns become more stable as the group studied is older Reading in between the lines of the most discussions of the interest test phenomena, this fact is taken to mean that Strong interest patterns are fixed, once developed, and therefore any actual changes are due to unreliability or other types of error But our theory can encompass the same phenomena without recourse to the catch-all concept of error. First of all, we assume that it would be acknowledged as a social-psychological and sociological fact that the older the individual is, the more likely it is that he will have established himself occupationally and the less likely it is that conditions will require a change in his occupation *In answering a Strong Vocational Interest test an individual is expressing his acceptance of a particular view or concept of himself in terms of occupational stereotypes*" (111 59 and 53)

Bordin therefore agrees with Carter in thinking of inventoried interests

as the reflection of a self-concept, which is developed as a result of the interplay between the endowments of the individual and the environment in which he lives He differentiates between Carter's view, which he considers dynamic because of this emphasis on interaction, and Darley's viewpoint, which he considers static because of its emphasis on the maturation of personality traits and their biological basis He characterizes Strong's viewpoint, which, like Darley's, he judged from personal conversations supplementing their writings, as empirical and going no further than stating that there are interest patterns which differentiate men in one occupation from those in others Reconciling the data on the stability of interests with Carter's theory of their dynamic nature by the use of Carter's interpretation of interests as self-concepts, he by-passes the first two objections just raised by this writer to Carter's, Darley's, and Berdie's viewpoints He states "If under personality we include the specific long- and short-term goal-directed strivings of the individual, then this view of interest patterns may be described as considering these patterns as by-products of the individual's personality We must recognize that these strivings are in a state of flux, changing to meet the fluctuations in the situation" (111 54) Bordin goes on to set up a series of challenging hypotheses and corollaries which he believes research will prove valid As most of them are still to be tested, they remain in the realm of hypotheses, in the writer's opinion, they seem sound He would, however, incorporate them in a concept of interest which puts less exclusive emphasis on personality and on environment, for the facts which have been reviewed justify assigning important roles also to ability and to heredity

As this introductory section on the nature and development of interests has taken on the proportion of a small monograph, thanks to the newness of the important work on interests, it may be well briefly to summarize the results of the research which have been reviewed in order to bring them into sharper focus before setting forth a theory of interest which they seem to justify

In summary, the inventoried interests of fathers and sons resemble each other about as much as do those of fraternal twins, whereas those of identical twins are considerably more alike, suggesting, since fraternal-twin environments are more similar to each other than are father-son environments, that heredity plays a part in the development of interests Interest patterns are related to degree of general intelligence, apparently because without understanding there can be no genuine and enduring interest or because a self-concept cannot endure unless it can be in part made a real-

ity There is no satisfactory evidence as yet concerning special aptitudes and interests Attitudes such as liberalism and social adjustment are related to interest patterns, even prior to occupational experience This is true also of values, which are presumably more deep-seated aspects of personality Personality adjustment in the sense of feelings of adequacy and security has not been shown to be related to interest patterns There is some evidence that temperament and endocrine make-up may be related to interest patterns, at least insofar as they affect masculinity and feminity, but the experiments in question are limited in number and in scope

Experiences such as courses in school and college, and staying in an occupation over a long period of time, have no effect on inventoried interests, although the experiences of the first two years of college training in a professional field have been shown, perhaps because of the importance of the first real contact with a field, to have some effect on inventoried interests Those who leave a field of training while in college tend to undergo a decline of interest in that field after leaving, and those who change to a field tend to show some increase in related interests after they have made the change, but these changes are not on the whole very great They are significant enough so that it is possible that some persons do show real changes of interests

A theory of interests which would take into account all of the above facts, without going beyond them, must recognize the significance of heredity, as shown in family resemblances and as implied in the data on aptitudes, personality, and endocrine factors, it must also recognize the role of experience, as shown in the data on modification of inventoried interests with change of type of experience An adequate theory of interests must build on the findings concerning the relationship between general aptitude and interest, which imply that in some instances aptitude probably does come first, resulting in approval, satisfaction, and interest. It seems probable that aptitude plays a part in the development of personality traits, as shown in certain studies of the effects of social skills on adjustment (395,583,498), and therefore in the development of interests as these are affected by personality. And it must recognize the fact that there are relationships between interests and the deeper layers of personality such as values and temperament, and possibly also personality traits and drives (although these last two relationships have not been and may perhaps not be established), and that these relationships are in some instances causal In other words, an objective theory would recognize the

fact of multiple causation, the principle of interaction, and the joint contributions of nature and nurture. It would read more or less as follows

Interests are the product of interaction between inherited aptitudes and endocrine factors, on the one hand, and opportunity and social evaluation on the other Some of the things a person does well bring him the satisfaction of mastery or the approval of his companions, and result in interests Some of the things his associates do appeal to him and, through identification, he patterns his actions and his interests after them, if he fits the pattern reasonably well he remains in it, but if not, he must seek another identification and develop another self-concept and interest pattern His mode of adjustment may cause him to seek certain satisfactions, but the means of achieving these satisfactions vary so much from one person, with one set of aptitudes and in one set of circumstances, to another person with other abilities and in another situation, that the prediction of interest patterns from modes of adjustment is hardly possible Because of the stability of the hereditary endowment and the relative stability of the social environment in which any given person is reared, interest patterns are generally rather stable, their stability is further increased by the multiplicity of opportunities for try-outs, identification, and social approval in the years before adolescence By adolescence most young people have had opportunities to explore social, linguistic, mathematical, technical and business activities to some extent, they have sought to identify with parents, other adults, and schoolmates, and have rejected some and accepted others of these identifications, self-concepts have begun to take definite form For these reasons interest patterns begin to crystallize by early adolescence, and the exploratory experiences of the adolescent years in most cases merely clarify and elaborate upon what has already begun to take shape Some persons experience significant changes during adolescence and early adulthood, but these are most often related to endocrine changes, and less often to changes in self-concept resulting from having attempted to live up to a misidentification and to fit into an inappropriate pattern Vocational interest patterns generally have a substantial degree of permanence at this stage for most persons, adolescent exploration is an awakening to something that is already there.

CHAPTER XVII

MEASURES OF INTERESTS

THE discussion of definitions at the beginning of the preceding chapter pointed out that the most productive work so far in the measurement of interests has been done with the *inventory* technique For this reason only one *test* of interests is considered in this chapter, although it is hoped that others will be sufficiently developed during the next few years to justify later inclusion This test is the *Michigan Vocabulary Profile Test*

As interest inventories have been developed in greater numbers there is, in that respect at least, a broader field from which to choose But it has apparently been much easier to write "like-indifferent-dislike" items than to ascertain what they measure and what their significance is for counseling and selection Some interest inventory authors have launched their instruments without validation data and have not followed them up sufficiently to make them useful Others, such as Garretson (279) and Dunlap (220,713) at the junior high school level, and Cleeton (161) at the adolescent and adult, have made careful and intensive studies prior to or immediately after publication, but have not followed through with further investigations of the nature of the traits measured by, or the validity of, their instruments Their inventories cannot therefore be considered as more than potentially useful tools One or two others, such as that by Lee and Thorpe (834), may in time be found useful, but data have yet to be made available to demonstrate their value. Any user of such an untried inventory in counseling or selection operates on faith alone—and faith is a poor substitute for facts in psychology and in occupations Two interest inventories and one values inventory have been studied over a period of years, and sufficient data have been accumulated to make them extremely valuable diagnostic instruments As is brought out in the discussion of the nature and development of interests, these are the *Strong Vocational Interest Blank*, the *Kuder Preference Record*, and the *Allport-Vernon Study of Values* The first-named inventory has been the subject of intensive study from many viewpoints over more than twenty years,

its author having assumed responsibility for integrating and interpreting the results of relevant research (775,798), the second was experimented with for several years before publication and has since been revised, and new studies by its author and others are continually appearing in the journals or in new editions of the manual (446,802), the last-named inventory has been used since 1933, during which time numerous psychologists have reported on it, and several have assumed responsibility for bringing these reports together and discussing the significance of the traits measured (136,216) These three inventories are therefore treated at some length in this chapter Much briefer treatments of the *Cleeton Vocational Interest Inventory,* and the *Lee-Thorpe Occupational Interest Inventory* are also included, as these are either widely used or new and well-publicized instruments, some of which include some novel features Also treated are trends observable in recent interest test construction, as this is a very active field and those who are experimenting with interest measures may find a brief discussion of some value, even though they are not presently useful to practitioners

Several studies have compared existing interest inventories in order to assess their relative value Some of these have used occupational criteria, effectively demonstrating the superiority of the Strong over the Hepner and Brainard inventories (84) But others have compared one test with another (e g. 301) and with occupational preference, thereby proving nothing unless one is willing to postulate the validity of one of the indices, the validity of which is in question

The Strong Vocational Interest Blank (Stanford University Press, 1927 and 1938)

The eminent group of applied psychologists assembled at the Carnegie Institute of Technology after World War I directed their attention partly to problems in the measurement of interests, particularly those which might differentiate salesmen from engineers The history of this work has been recounted by Fryer (277 Ch 3) and need not be repeated here, beyond stating that Strong began his work with the inventory technique as a member of this group, and took it with him to Stanford University, where Cowdery (174) and other students worked with him in establishing it as an effective method of differentiating between occupational groups Strong published his first edition of the blank in 1927, after several preliminary studies had shown the validity of the approach, a new revision that is currently in use was brought out in 1938, based on the work of

the intervening years, the many studies of the nature of the traits measured and of their validity in educational and vocational counseling and selection were brought together in his monograph of 1943 (775), new occupational scoring keys are added from time to time as studies are completed (Paterson is now revising the psychologist key, and Schwebel is developing one for pharmacists), and the journals continue to carry new studies of various aspects of the Blank's significance and use. It is without question one of the most thoroughly studied and understood psychological instruments in existence.

Applicability Strong's Vocational Interest Blank was developed for use with and standardized upon college students and adults employed in the professions and in business Because of this it includes some terms which are unfamiliar to high school students and to adults in lower level occupations For example, even high school juniors and seniors filling out the blank often ask the meaning of terms such as "sociology," "physiology," and "smokers" and reveal a complete unawareness of the nature or existence of the magazine *System* For these reasons the question of the use of Strong's Blank with persons of less than college level has frequently been raised It can be answered from two sets of data Strong's and Carter's studies of the interests of adolescents and adults, already discussed in some detail, and a recent investigation by Stefflre (752).

The age and stability studies, both cross-sectional and longitudinal, have been seen to show that meaningful data can be obtained by means of Strong's Blank from boys and girls as young as 14 or 15, and that by the time they are 18-to-20-year-olds their Strong scores are rather well fixed This suggests that, despite the apparent difficulty of some of the words used in the inventory, it is sufficiently well understood at those age levels to be applicable to most high school students

The vocabularies of the Strong, Kuder, and other inventories were analyzed by Stefflre, who reported that the Strong Blank has a 10th grade vocabulary This fits in with the data on its usefulness with 17-year-olds, and suggests that it should be used below that level only with the more able and more advanced students

Potential users of interest inventories often ask whether a subjective technique such as this is subject to faking when used in selection programs, and even in counseling, because of the desire to make high scores in some occupations. The job applicant wants to appear in the best possible light, and even if he is above conscious distortion there are many genuine opportunities to give oneself the benefit of the doubt in answer-

ing an inventory The student seeking guidance may be eager for self-insight and for an objective picture of himself, but in answering the questionnaire he is nonetheless guided by his self-concept, set of occupational stereotypes, and a desire to appear favorably in the eyes of the counselor Strong (775 684) and Steinmetz (754) have experimented with deliberate faking in students, first administering the inventory in the standard way, and at a subsequent session administering it with directions to attempt to raise the score on a specific occupation (engineering and school administrator, respectively) In both instances very great changes resulted, the mean scores shifting to such an extent that the majority received A ratings, in contrast with B+'s (engineers as engineers) and C's (business students as engineers, education students as administrators) Other scores were affected by these distortions, as would be expected in view of the intercorrelations

Faking by job applicants, a much more important experiment than deliberate faking by students, was also checked by Strong (775 688–690) who administered the Blank to 118 men responding to an advertisement which he inserted in newspapers The inventory was given as a preliminary hurdle for life insurance sales positions, some, according to Strong, took the questionnaire out of mere curiosity, but an indeterminate number of others were more serious in their purpose. The scores made by these men in their then occupations were compared with their scores as life insurance salesmen, with the finding that only groups whose averages were above a standard score of 40 on the sales key were already employed in some kind of sales work The conclusion was that, although some individuals may have intentionally raised their scores somewhat, the majority did not achieve, or perhaps even try, any appreciable distortion According to Strong (775 688), Bills did find that applicants under 24 years of age who scored A on both sales keys were less likely to succeed as insurance salesmen than those who scored B+, or B+ and A, on the two scales, presumably because of bluffing In selection, therefore, the use of other checks on interest inventories is probably desirable

As for counselees, there is no experimental evidence that their scores are or are not affected by the desire to appear, to themselves or to the counselor, in a certain light While Spencer (733) has shown that some personality inventory items are answered differently when a name is signed than when answered anonymously, he also showed that answers to other items, the least personal and the most like those in interest inventories, are not changed because the respondent can be identified

Conscious distortion in counselees or students can therefore probably be dismissed as negligible No one has as yet found a way of checking up on unconscious distortion, although it might be tried under hypnosis or narcosis

Content The Vocational Interest Blank Form M (men) consists in its present form of 400 items grouped according to type of content The first group is a list of many types of occupations at and above the skilled level, emphasizing the business and professional fields This is followed by lists of school subjects, amusements (games, magazines, sports, etc), activities (hobbies, pastimes, etc), peculiarities of people, vocational activities, factors affecting vocational satisfaction, well-known persons exemplifying occupational stereotypes, offices in clubs, and ratings of abilities and personality characteristics (the actual grouping is not quite as in this list, which is based on content rather than on form of item). The women's form has 263 items in common with the men's and a total of 400 in the revised form

Administration and Scoring There is no time limit for the Vocational Interest Blank, as the task is to answer all questions, the time required ranges from a little over 30 minutes for superior, well-adjusted, adults to something more than an hour for less able or less stable individuals It is well to allow an hour when testing groups, and to administer the blank at the end of a test session, e g , just before a rest period, when it is part of a battery This makes it possible to dismiss subjects as they finish, but does not put too much pressure on those who have not finished In guidance centers the inventory is often given to older adolescents and adults to complete on their own time at home, this works well when the client has a place to work without having his responses affected by the comments of on-lookers, and when he understands the importance for himself of filling it out rapidly and without consultation.

In answering the blank, the subject marks each item according to whether he likes, dislikes, or is indifferent to it. The answer to each item is assigned a weight based on the degree to which the answers of men in a given occupation, e g engineering, differ from those of men-in-general This procedure is sufficiently different from those normally used in developing scoring procedures to be worth describing, for understanding it means practically an understanding of Strong's Blank. Table 28 presents the Strong's data for one item, "Actor," showing the responses of engineers and of men-in-general.

It is made clear by the "difference" row in Table 28 that engineers are

TABLE 28

DETERMINATION OF WEIGHTS IN STRONG'S BLANK ITEM "ACTOR"

Group	*% Like*	*% Indifferent*	*% Dislike*
Engineers	9	31	60
Men (Gen'l)	21	32	47
Difference	−12	−1	13
Weight	−1	0	1

less likely to indicate a liking for the occupation "actor" than are men-in-general, slightly more likely to indicate indifference, and much more likely to show a disliking for it By means of a formula based on the significance of the difference between two percents these data are converted into the weights shown in the bottom row In scoring the inventory of a young man who thinks he wants to be an engineer, but who indicates that he would like being an actor, one would therefore deduct one point from his engineering score he has shown that, in this respect at least, he is more like other men than like engineers It is perhaps worth noting that this is true, even though other men tend not to like being an actor, for they indicate a liking for it more often than do engineers

The score for engineer, then, is the algebraic sum of the weights corresponding to each answer marked by the client, a total of 400 weights A comparable addition and subtraction must be made for every occupational or other score (e g , masculinity-feminity) desired by the counselor To do this by hand is time-consuming, for it takes a novice about 15 minutes to score one blank for one occupation even with the stencils provided for this purpose, and the men's inventory is scored for more than 40, the women's for more than 20, occupations and traits With the aid of two Veeder counters (Nos ZD-18-T and ZD-8-T, Veeder Mfg Co , Hartford, Conn) an experienced scorer can cut the time in half, averaging about ten occupational scores per hour As this would still mean about four hours scoring time per men's blank when all keys are used, machine scoring is necessary when any number of subjects and occupational scales are involved.

Strong describes the methods in his manual Briefly, they are the Hollerith machine, reading the answers from the Blank, at a cost of about $1 25 per blank for 39 men's occupations (the price varies), the IBM method, in which a special answer sheet and electrographic pencil are used (as with most standard tests for machine-scoring), at a similar cost for all men's scales, and the Hankes method (Engineers Northwest, 100 Metropolitan Life Bldg , Minneapolis 1, Minn), requiring a Hankes answer sheet, at 70 cents per blank for all 42 current men's scales or all

24 current women's scales The names and addresses of organizations having scoring machines and offering scoring services to others are listed in Strong's manual, which is kept up to date, the Hankes' method is described in a paper by Strong and Hankes (782)

The cost of scoring Strong's Blank has been something of a deterrent to its use in some institutions, more often in public schools than elsewhere, and the more recently published inventories with their less expensive scoring have for this reason had a wide appeal Many a user has bought them frankly as less expensive substitutes for Strong's Blank Because of the pressure to cut down costs, Strong and many others have attempted to simplify the scoring as much as possible, these have been summarized in Strong's book (775· Ch 24) and followed up by another study (778) Only Strong's conclusions can be cited here weighted scores differentiate better than the unit scores proposed by Dunlap and others and should therefore be used in counseling and selection Weighting each item one, instead of from −4 to 4, would, Strong has shown (778), lead to different counseling in from one in every twelve to one in every six cases When the cost is approximately one dollar per case the price of greater validity does not seem unduly high Public schools and other institutions spend far more per pupil on things of less significance than finding out what kinds of educational and vocational activities are most likely to challenge them As a compromise, Strong has devised six group scales, one each for the biological science, physical science, social welfare, business detail, contact and linguistic groups These correlate fairly well with the specific keys and can be used when only directional counseling is needed

Scores on Strong's Blank are recorded in many different ways by users of the inventory One frequently sees reports in which the occupational scores are arranged in order of magnitude, all occupations in which A's are made being grouped first, the B+'s next, and so on This is done on the assumption that the counselor and client are most interested in the A scores or, in their absence, in the B+'s This method has two drawbacks it focuses attention on specific occupations, and it makes it difficult to perceive patterns of scores Each of these is worthy of brief discussion

The Vocational Interest Blank can be scored for about 40 occupations, and the number may conceivably be increased to 45 or 50 in due course. But there are nearly 30,000 jobs in the *Dictionary of Occupational Titles* (888), and, while many of these are more specific than those in Strong's Blank, and could be combined to make a smaller number, it would still

be true that interest in most occupations cannot be scored on Strong's Blank It is manifestly unwise, then, to play up scores on specific occupations. The result too often is that a student says, "I rate A as a minister, but I don't have any desire to be a minister," and the insights into interests which might be gained from that score are lost in the negative reaction to a stereotype of a specific field, or a client leaves the counselor and reports to his family that "One test showed that I should be a personnel director I wonder what the boss will think of that?" missing the more general implications of that high score

When occupational interest scores are grouped according to their factorial composition, however, the result is often quite different This puts related occupations together in families, it permits the analysis of scores in terms of types of occupations rather than specific occupations, and it makes it easy to see whether or not a high score in one occupation is supported by high scores in related occupations Thus an A as physician, for example, is a much surer basis for encouragement in choosing a premedical or biological sciences major if supported by A's or B+'s as psychologist, dentist, chemist, and engineer than if the scores of these occupations are largely B+'s and C's. The report sheet published by Strong, the Hankes Report Form, and many others are organized in such a way as to make possible this type of pattern analysis (see Fig 7)

Pattern analysis was first described in some detail in a booklet by Darley (189 Ch 2), in which a helpful distinction is made between *primary, secondary,* and *tertiary* interest patterns He defines as primary interest patterns those fields in which the letter ratings received are largely A's and B+'s, as secondary patterns those occupational families in which scores are predominantly B+ and B, and as tertiary those in which they tend to be B's and B−'s Using this classification of the letter ratings received by a counselee makes it possible to focus attention on the kinds of occupations which he is likely to find congenial It is more helpful to know, for example, that his primary interest patterns are in the scientific and literary occupations with a secondary pattern in the social welfare field, than to know that he made A's as psychologist, physician, physicist, chemist, engineer, personnel director, public administrator, advertising man, author-journalist and president of a manufacturing concern Darley tabulated the frequency of interest types or patterns for 1000 men at the University of Minnesota, it is worth noting that approximately half of these men had no primary interest pattern This is discussed at great length in connection with the use of Strong's Blank

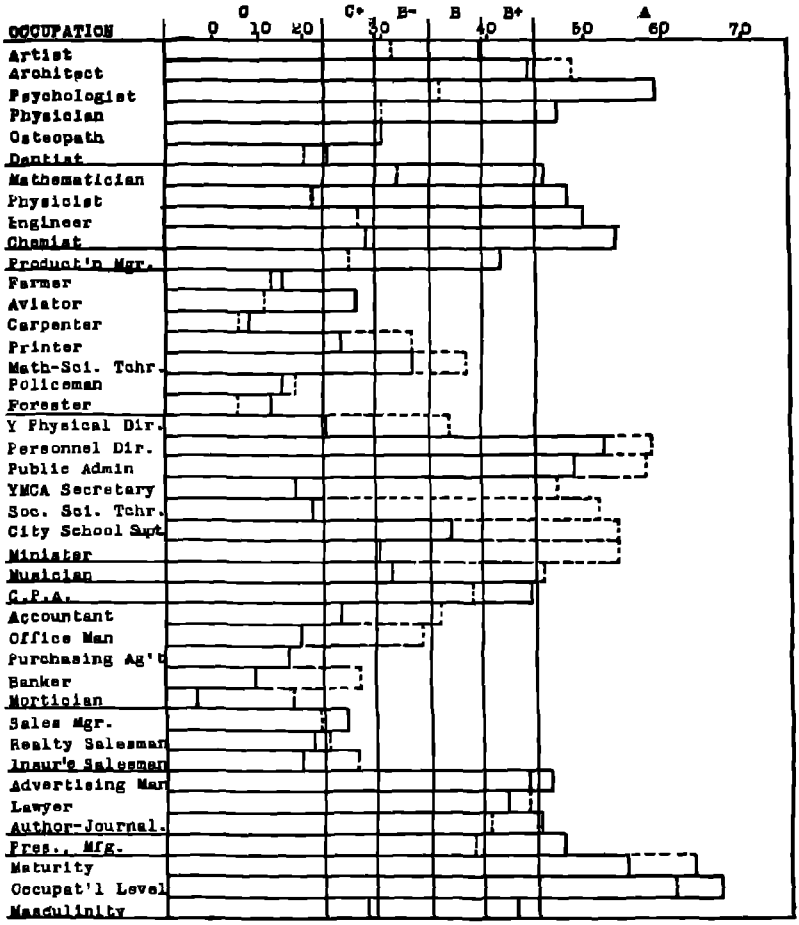


FIGURE 7

INTEREST PROFILE OF A YOUNG MAN AT AGE 25, SOCIAL STUDIES INSTRUCTOR, AND AT AGE 37, AS PSYCHOLOGIST (STRONG'S BLANK)
(Broken lines at 25, solid lines at 37)

It should also be noted that, as might be anticipated, the use of interest patterns has been found more valid, with entry into an occupation as criterion, than specific occupational score (926)

Norms The question of norms is in fact a double-barrelled question, for it concerns both type and number of cases As the details of both of these are given in Strong's manual and in his book (775 694–702) they

need not be reproduced here, but one frequently encounters mis-statements made by presumably well-informed users of Strong's Blank. For example, a specialist in the selection and training of nurses once stated that the nursing scale of the women's form was of little value because it was based on about 100 nurses from one hospital in Chicago, it was actually based on approximately 400 nurses, 283 located in, but not necessarily natives of, the nurse-importing city of New York, the other 117 from upstate New York and elsewhere This is not a balanced sample, but neither is it as unbalanced as the above-quoted critic implied, as data on the validity of the women's form will later make clear, it is also not the reason for the lack of correlation between scores and grades in nursing schools Some generalized statements concerning the norms of the Vocational Interest Blank follow, in order to provide the orientation to the base of this inventory which many users apparently lack

The data concerning numbers are simple enough Cowdery's early work (174) showed that occupational differentiation could be achieved with groups of as few as 35 persons For this reason, Strong's first keys were based on about 150 cases each, surely a conservative application of Cowdery's findings Subsequent work (775 639–650), however, led him to increase the number in order to increase the discriminating power of the test, the numbers were therefore raised first to 250 cases per occupation, and then to between 400 and 500 cases Accordingly, the earliest scales are based on groups of from 150 to 200 cases per occupation, the newest scales on between 400 and 500 persons Evidence reported by Strong shows that these numbers are large enough to minimize shrinkage of mean scores in cross-validation

The question of type or quality is more complex, and can itself be broken down into several questions Outstanding among these are, 1) the criterion of success which warrants inclusion of a given case in the criterion group, 2), the representativeness of the sample, and, 3), the timelessness of the sample or the degree to which the interests of successful psychologists of 1928 are representative of the interests of successful psychologists in 1948

The criterion of success varies from one occupation to another, as one might expect in view of the differences in occupations output may measure success in electrical unit assembly work, but not in teaching Strong's life insurance salesmen sold at least $100,000 worth of insurance annually for three years. In the cases of occupations which have some accrediting or other evaluative procedure of their own it was used

architects were members of the state board of architecture, carpenters were union members, certified public accountants were certified in their states, chemists were non-professorial members of the American Chemical Society, and psychologists were Fellows (full members) of the American Psychological Association When no such criterion of being established in a field was available, other evidence of status was used the journalists were editors listed in *Editor and Publisher Yearbook,* city school superintendents were employed in cities of more than 10,000 inhabitants, personnel managers were "carefully selected by competent authorities," and YMCA physical directors were selected by a YMCA college All members of criterion groups had been employed in their occupation for at least three years, and none were over 60 years of age In some cases the criterion was probably not as stringent as in others the apparently miscellaneous collection of office workers were probably not as highly selected, in their field, as the physicians who graduated from Yale and Stanford were in theirs, especially when it is considered that the great majority of the physicians practiced in favored areas Sometimes the criterion was established and the group selected by Strong (e g , psychologists) and sometimes it was others who did these things (e g , men teachers) On the whole, however, there is little to quarrel with in the criterion.

The representativeness of the sample is more difficult to judge, and has not been investigated sufficiently to provide answers to the serious questions which can be raised concerning some of the groups The purchasing agents were located in Northern California, Los Angeles, Washington, D C , and Cleveland, the psychologists were scattered throughout the United States and constituted 35 percent of the population from which they were drawn, the personnel managers came from New England, the Middle Atlantic States, the Great Lakes, and the Pacific Coast, and the city school superintendents were located in various parts of the States these seem reasonably likely to prove representative But the male social-science teachers were all from Minnesota, and may have had interests quite different from those of their confreres in Vermont and Alabama, the real estate salesmen were all from California, and may differ considerably from those of Massachusetts, though no doubt very much like those of Florida (!), and the farmers were all from the West Coast, and perhaps unlike those of Maine and Georgia No check has been made of the possible existence of regional differences in occupational interest patterns, nor even of the existence of regional differences in the interests

of men-in-general Strong does report an unpublished study by Pallister and Pierce (775 674–677) which compares the interests of Scotsmen with those of Americans, the former were artists, journalists, ministers, and policemen living in or near Dundee The interests of Scottish artists and policemen were very much like those of their American counterparts, while those of ministers (possibly) and journalists (clearly) were different. Strong concludes that the differences cannot be attributed to language usage, since at most two occupational groups differ, he believes that they are due to differences in sampling, and points out that the American journalists were a highly selected group (listed in *Who's Who,* etc) while the Scots constituted all the literary employees of one local publishing house There is also the possibility of national, and therefore regional, differences in the selection of persons in some occupations. Second-generation Japanese high school boys, born in America, were found to resemble white Americans in their interests in another study reported by Strong (775 677–679), leading him, as in the case of the Scottish study, to conclude that richness of meaning has little effect upon responses to interest items, and that the omission of some terms which are not understood does not appreciably affect interest inventory scores Although they have only indirect bearing on the question of regional differences in vocational interest patterns in the United States, these findings do indicate that the latter may not be as important as *a priori* reasoning might suggest It is to be hoped that investigations of the differences between social studies teachers in the Midwest and East, ministers in the Middle Atlantic and Southern States, and other regional occupational groups will in due course be made.

The temporal validity of Strong's occupational interest scales, like their regional validity, has not been the subject of published investigations Professional self-consciousness and the rapid development of their own profession have, however, made psychologists conscious of the problem It is frequently pointed out, for example, that Fellows of the American Psychological Association in 1928 were largely laboratory psychologists, more interested in problems of mental organization and functioning as shown in introspective or experimental studies of learning in humans and in animals than in problems of human adjustment, and that, in contrast, the tremendous growth of industrial, educational, and clinical psychology in recent years now puts the heirs of these theoretical psychologists in the minority The interests of the two generations of psychologists may be quite different, or, on the other hand, the common

core of interest in the scientific study of man may be so important that, when compared with other professional men, they may seem quite similar As mentioned in the first section of this chapter, Paterson is conducting a study which will throw light on the problem of interests in a changing profession If he reports little change, other keys may be used with some confidence, for psychology appears to have been changing more than most occupations, if he does find differences, caution will be needed in the use of scales for other occupations which may have changed Inspection of the list suggests only one other which may have been affected in a comparable way, that of YMCA secretary, in which profession the emphasis seems to have changed from personal-religious to social

Another aspect of the norming of the inventory which needs consideration is the *form in which its scores are expressed* and the normative group to which it compares a person Strong provides distributions of raw scores based on the appropriate criterion group (e g , engineers), standard scores and letter grades for the same groups, and percentile scores for the criterion group, Stanford freshmen, and Stanford seniors This plethora of norms raises the question of which to use As was pointed out in an earlier chapter, general norms such as those of the student groups have little value, for they tell nothing of the individual's prospects of success in competition with selected occupational groups It is therefore the norms for the criterion group which should be used As is pointed out in the manual, the letter ratings have the advantage over percentiles and standard scores in that they indicate clearly and readily whether or not a person's interests resemble those of men or women successful in the occupation in question, without obscuring the issue with the problem of little understood differences of degree For although the difference between the 60th and 90th percentiles on an aptitude test has known significance, that between the same percentiles on an interest inventory such as Strong's does not there is no reason for thinking that a *high* degree of resemblance to the interests of the *average* successful worker is superior to a moderate degree of resemblance The man at the 60th percentile might actually differ from the average successful worker in ways which make him more like the most successful or satisfied workers than the man whose 90th percentile rank indicates closer resemblance to the average established man In both counseling and selection, therefore, it is better to use the letter ratings.

These are so established that the top 69 percent of workers in the

occupation are assigned scores of A, and the bottom 2 percent are assigned scores of C Thus anyone resembling the majority of established workers is assigned an A or at worst a B+, and all persons who rate C on an occupation are quite unlike the bulk of men in the field in question Scaled scores may be useful in certain types of studies, as when differences between groups are being studied

Standardization and Initial Validation Much of what might normally be discussed under this heading has already been treated under the sections on scoring and norms, because of the unique nature of the Vocational Interest Blank as an inventory based on group differences and scored by different keys for each occupation on which it has been standardized It is also difficult, in a sense, to distinguish between initial validation and subsequent validation, because of the basic nature of some of the later studies which Strong has made of his inventory However, it will clarify matters briefly to outline the steps gone through in the first validation of each occupational scale of the Strong Blank

The Blank itself, it will be remembered, was the result of several years of experimentation by Strong and his associates and students his list of 420 items, later abbreviated to 400, consisted of those which had been found most useful in these various studies In devising the scoring scale for each occupation the inventory was administered, often by mail, to men or women who had been in the occupation in question for at least three years and who, in most cases, were distinguished by having been nominated by well-informed persons as leaders in their fields, by being listed in the appropriate *Who's Who,* or by professional certification. The scores made by these 200 to 500 persons (see section on *Scoring* for methodology) were distributed on the normal curve and converted into standard scores and letter grades It should be noted that in this procedure the norm group consists of the same persons who constituted the criterion group, and that experience has repeatedly shown that when these two groups are the same, the mean scores of subsequent groups will be lower than those of the norm, even though all groups are random samples of the same population. This has been noted and pointed out by Strong (775 649,675), in experiments which showed that, when the criterion-norm group consists of 250 persons, the shrinkage of scores will be about 1 50 standard scores He felt that it was wise to continue this procedure, however, in order to have the largest possible criterion groups This was justified by the fact that the shrinkage for a criterion group of 300 was only 0 90 standard scores As there was very little change for

numbers above 300, Strong's choice of criterion groups of between 400 and 500 seems wise, but users of the blank must allow for a shrinkage of about one standard score—not enough to be important in most individual instances, but at times making the difference between a B and a B+, a B+ and an A The experience of psychologists during World War II, when, for example, norms were regularly gathered for 400 aviation cadets per day in one center, brought home the need for cross-validation as a means of avoiding shrinkage even supposedly similar groups of 1000 cases frequently showed significant differences It is therefore to be regretted that Strong did not correct his norms by the amount appropriate to the number of cases used, or subsequently obtain data from new normative groups.

The procedure described above is the standard method of developing an occupational scale for Strong's Blank Many other studies have been conducted which validate the inventory, either through cross-validation or by other means we have, for example, seen studies of the validity of responses (faking), age changes, and the effect of experience Other studies have considered the relationship between interest-inventory scores and grades or sales production, but as these fit more naturally into the discussion of field validation they will be taken up after the next section.

Reliability The odds-evens reliability coefficients of 36 of the revised men's scales are reported in the manual as averaging 88, based on the records of 285 Stanford seniors, only one coefficient was below 80, the reliability of the CPA scale being 73 Taylor (813) found retest reliabilities averaging 87 for high school boys and 88 for girls, on the appropriate forms The retest reliability was ascertained for college students by Burnham (125) with eight of the original scales, the average being 87 Strong obtained retest reliabilities after five years, the first testing being when the 285 men in question were college seniors these averaged 75, and must be thought of as not only an index of the reliability of the scales, but also as a measure of the stability of interests in early adulthood For 10th graders retested after two years the mean for 7 typical scales was 57 (135), for 11th graders after three years it was .71 (813). It is evident that the scales are reliable enough for confident use in individual diagnosis at least after age 17

Validity The validity of Strong's Vocational Interest Blank has been investigated by relating the scores of its various scales to those of other tests, to grades in school and college, to completion of training, to earnings in sales work, to ratings of success in various types of work, to per-

sistence in an occupation, to differences between occupational groups, and to job satisfaction As this suggests, there has been accumulated an unusual amount of validation data, even for an instrument which has been in existence for twenty years

To attempt to review all of these validation studies would be not only a sizable task [Strong's monograph (775) attained 746 pages even after whole sections had been ruthlessly cut], but, because of Strong's thoroughness, an unnecessary one There are, however, two reasons for discussing certain selected studies here at some length 1), an understanding of these details is essential to adequate use and interpretation of Strong's Blank, and, 2), they should become an integral part of the literature on vocational tests in order that the first objective may be attained Some of the studies discussed by Strong are therefore treated here, together with others of special significance which have appeared since he compiled his review

Tests of intelligence have been correlated with Strong's scales in eight studies summarized by Strong (775 333–334) and in others subsequently published (776) The various investigators agree that the relationships with scientific and linguistic interests are positive, the former being moderate or low but significant and the latter so low as to be of little meaning, as shown in Table 29

The correlations with social welfare and business interests tend to be

TABLE 29

RELATIONSHIP BETWEEN INTELLIGENCE AND INTERESTS
(Taken from Strong, Table 90)

Occupational Scale	*Correlations*					
Psychologist	37	43	41	36	15	38
Physician	16	27	19	04	10	24
Engineer	21	20	14	17	08	28
Chemist	30	34	15	31	03	35
Advertising Man	02	14	12	− 11	45	01
Lawyer	07	21	20	13	39	13
YMCA Secretary	− 22	− 19	− 18	14	− 15	− 18
Personnel Manager	− 16	− 10	− 13	27	− 07	− 02
City School Superintendent	− 12	03	01	32	06	− 06
Office Worker	− 31	− 27	− 28	09	− 38	− 25
Purchasing Agent	− 25	− 33	− 31	00	− 07	− 21
Life Insurance Sales	− 35	− 34	− 31	− 19	00	− 26
Vacuum Cleaner Sales	− 36	− 40	− 40	− 14	− 36	—

negative, although most of the coefficients are so low as to make them of little practical significance despite their theoretical implications Typical data are reproduced in the lower half of Table 29 It will be noted that, although there are occasional discrepancies, there is sufficient agreement so that analyzing the nature of the groups and of the intelligence tests used in an attempt to reconcile them is unnecessary

Special aptitudes have not been correlated with scores on Strong's Blank in many published studies. The relationship between the Stanford Scientific Aptitude Test and Strong's six group scales was ascertained by Long (478) for 200 college students although, as he points out, it is not at all certain what the former test measures He found significant positive correlations (.26 and .50) with Strong's two scientific scales, and a significant negative relationship with the business-contact scale (−.37), the others were negligible, and none could be explained on the basis of intellectual differences as measured by the A C E Psychological Examination Leffel (460) correlated scores on the O'Rourke Mechanical Aptitude Test with Strong scores, showing positive relationships (.42 and .46) between the O'Rourke and the keys for chemist and engineer, and negative relationships (−.25 and −.25) with the scales for social studies teacher and lawyer Holcomb and Laslett (375) found comparable results for the Stenquist Mechanical Aptitude Test This suggests that aptitude, being more fundamental than interest, may have some causal effect on the latter, but as noted previously the O'Rourke and the Stenquist are information tests the scores of which are no doubt influenced by both aptitude and interest, making it impossible to infer causal connections The latter study also used the MacQuarrie, which correlated .22 with Strong's engineer scale Moore's study (536) showed correlations of .30 and .35 between the Bennett Mechanical Comprehension Test and the engineer key, and of .21 and .26 with the aviator scale, while the correlations between Bennett and Strong production manager and carpenter scales were negligible As the MacQuarrie and Bennett tests are more strictly measures of aptitude than the O'Rourke and Stenquist, it may perhaps be inferred that aptitude plays a part in the development of interest This seems warranted despite Klugman's (435) contrary finding with the Minnesota Clerical Test and Strong's women's clerical keys clerical interests appear to have too little significance in women to attach importance to findings based on them (see p 436)

Interest and values inventories with which the Strong Blank has been correlated include the Allport-Vernon Study of Values and the Kuder

Preference Record Data from three studies of college students, one of men (667) and two of women (188,217) show that the relationship between biological science interests and theoretical and aesthetic values is positive, while that with economic and political values is negative, the relationship is positive between physical science interests and theoretical values, and negative between these interests and political values, social welfare interests and social and religious values positive, business contact interests and economic and political values positive, the theoretical values negative, literary interests and aesthetic and theoretical values positive, economic negative, and business detail interests and economic and political values positive, theoretical negative The theoretical significance of these relationships was discussed in connection with the nature of interests,

As the Strong Blank is the better understood of the two interest inventories, the discussion of its relationship with the Kuder Preference Record will be postponed until that instrument is the focus of attention

Personality inventory scores have been related to Strong's scales with results which are somewhat contradictory. These studies are discussed in the preceding chapter and, as there is little in the way of generalizations to be drawn from them which is of value in using Strong's Blank, they are not summarized here

Grades and scores on achievement examinations have frequently been correlated with scores on Strong's inventory in the hope that the prediction of educational success would thus be improved The predictive value of scholastic aptitude tests being far from perfect, it was reasoned that motivation might account for part of the discrepancy, and the motivation and interest should overlap to some extent Accordingly, a number of studies were made, many of which were not published and only a few of which are cited here Townsend (851) ascertained the relationships between Strong's scales and scores on objective tests of school achievement made by groups of 50 to 100 boys in private secondary schools, and reported that they were few and significant only in the case of mathematics-science teacher and chemistry ($r = .36$), accountant-chemistry (.49), CPA-chemistry (.42), and mathematician-geometry (.31) Achievement in English and history were not related to social science teacher or author-journalist interest scales A procedure used by Segel (702) is suggestive, for after correlating Strong's scales with Iowa High School Content Examination Scores and obtaining correlations between scientific scales and scientific subjects which ranged from .28 to .49, but which were

not significantly related in other expected ways except for some negative relationships, he proceeded to use differential achievement scores. These consisted of the differences in the scores of two achievement tests, for example, the difference between literature and science, which had a correlation of 25 with life insurance sales interest The correlations, both positive and negative, were generally higher, those for scientific interests and differential scientific achievement (e g, science minus history and social science scores) ranged from 29 to 57 Similar relationships were found when school grades were used instead of achievement test scores in this study, although the trends were not so clear cut, presumably because of the more numerous other factors which affect grades The reason for the relationships between differential achievement and interests being greater than the relationship between achievement and interests is that, in the former, the effects of general ability are held constant and those of differential motivation and application are emphasized If a student consistently makes B+ in one field, and A in another, the relationships with interests will not be clear, but if his relative superiority in the second subject is brought out and correlated with interests, and the relative inferiority of his performance in the former subject is similarly treated, the role of interest is more likely to manifest itself

Grades in college were related to scores on Strong's scales by Alteneder (14), who worked with freshmen at New York University. The relationships were low (r's ranged from — 28 to 30), but she reported that low scholarship men tended to make higher engineering interest scores than high scholarship men, who tended to have interests more like those of teachers and CPA's, while low scholarship women had interests somewhat more like those of insurance saleswomen and stenographers than high scholarship women, whose scores as librarians, social workers, and lawyers tended to be high

Typing and stenography grades of about 100 women liberal arts college students were studied with the women's stenographer scale by Barrett (46) at Hunter College She reported only the data from tests which showed some validity, as the Strong scale "failed to show any significant relationship to grades" data concerning it were not reported

Engineering grades were correlated with Strong's engineering scale by Berdie (78), Campbell (775 521), and Holcomb and Laslett (375). In the first study the honor-point ratios of 154 University of Minnesota students were the criterion, their correlation with interest scores was .13 It is

worth noting that having a variety of interests, rather than only scientific, had no detrimental effect on achievement In the second study the correlation was .32 In Campbell's study of 270 engineering students at Stanford it was .185 For this same group the correlation between social-science interests and grade-point ratios in social science was .31 (social studies teacher). Holcomb and Laslett reported a similar correlation (.32) between engineering interest and engineering grades

Dental grades were used as a criterion for the dentist scale by Robinson and Bellows (634), who found a significant relationship (r = .19, .18, .19) Data on 141 dental students were reported by Strong (775 523), who found that those rating C on his scale made inferior grades (grade-point ratio of 2.01), while those of others were slightly higher (2.41 to 2.58) The significance of the difference is not reported

Medical school grades were subjected to study by Douglass (205) and by Jacobson (397), the former reporting that Strong's Blank was not useful in predicting success, the latter, however, finding that the first-year grades of students who were characterized by scientific and other interests were better than those with other interest patterns, those with medical interests as their only strong scientific interest but with other types supplementing these ranked second, and those with no scientific interests ranked at the bottom In connection with the top-ranking, broad-interest, students, it is interesting that Berdie (76) found that those with many "likes" get better grades than those with few "likes"

Teachers' college students were the subjects of studies by Goodfellow (295), Mather (775 526), and Seagoe (668) Goodfellow found, as Strong did with dental students, that those who rated A on the appropriate scale made better grades than those who rated C, the differences being significant Mather and Seagoe, however, both found no relationships between grades and interests

These contradictory findings in different studies of the relationship between interests and achievement, reported even for the same subject-matter or professional fields, might be explained on the basis of the unreliability of the criteria in some studies, the limited range of interests in most schools, and perhaps other factors which vary from one institution to another It is interesting that in none of the published studies has the criterion been subjected to any scrutiny, either as to distribution of grades or as to reliability, although in numerous studies (e g , 929) it has been demonstrated that the apparent lack of validity of the predictor is attributable in the first place to lack of reliability in the criterion The

limited range of scores in professional-student groups has been commented on by Strong (775 525–526), who contrasted the percentages of dental students receiving the various letter ratings on his scale with the percentages of college students in general the latter received fewer A's and more C's This phenomenon suggests the need to use an approach other than the correlational in studying the relationship between interests and educational achievement, and perhaps a criterion other than grades Several studies have used another approach, but before describing their results attention should be focused on one study of interests and grades in which the range of the former was relatively great

Personnel-psychology students in the Army Specialized Training Program, 95 in all, were the subjects of a study by Strong (776), after the publication of his book and the expression of opinions which might have been somewhat modified had this study been completed first In this investigation the correlation between intelligence and grades in psychology courses was only 20, whereas that for psychologist interest was 275 Neither of these is quite significant (31 required), but when individual course grades are considered the picture is clearer correlations for testing and social psychology courses were 355 and .150 for intelligence, and 32 and 34 for psychologist interests As the tendency in other studies has been for intelligence tests to be considerably better predictors than interest inventories, possible reasons were investigated It was found that, since the soldiers in question had all been selected partly on the basis of scholastic aptitude (minimum score of 115 on the AGCT or in the top quarter of white soldiers), the range of intelligence in the group was restricted in fact, 90 of the 95 made scores of 120 or above In the typical college freshman class, however, the tail of the distribution does not end so abruptly (see Chapter 6) The range of interest in psychology was, however, considerably greater, from low C to A, with a mean of low C Some of the men in the class reported, in conversation with Strong, that they had been assigned to personnel-psychology training without being consulted This is quite different from the typical college situation, in which the student is more generally in college for some reason of his own, and has something to say about the curriculum in which he studies Even the required courses are then in one sense electives, something accepted because they lead to some desired goal, be it no more than playing football or being with friends In the typical college class a student can therefore make the grade if he has the ability, regardless of interest in the subject-matter of the course, in the ASTP many students

lacked the motivation to use their ability In such circumstances one would expect to find, as Strong did, that interest has approximately as high a correlation with achievement as does aptitude

Completion of training is the other criterion of educational achievement used by some researchers It avoids the fine distinctions which grades attempt to make, stressing the more carefully considered and perhaps more clear-cut distinctions between, 1), passing and failing, and, 2), liking and not liking In Goodfellow's study (295), for example, it was noted that the education students who changed to other curricula made lower scores than did those who remained in education, and Strong (775 524) found that only 25 percent of the dental students who rated C on his dentist scale graduated in from four to six years, whereas 91 percent of those who rated A, 93 percent of those rating B+, and 67 percent, each, of those rating B and B−, graduated These findings fit in with Strong's explanation of the role of interest in educational achievement (775 529)

"If a student has sufficient interest to elect a course, his grade will depend far more on his intelligence, industry, and previous preparation than on his interest Interest affects the situation, however, in causing the student to elect what he is interested in and not to elect courses in which he is not interested When a student discovers he has mistakenly elected a course in which he has little interest, he will finish it about as well as other courses but he will not elect further courses of a similar nature"

To this should be added, in view of the one study in which the range of interest was adequate When a student is compelled to take a course or to study in a field not of his own choosing, the relationship between interests and achievement will be more nearly comparable to that of intelligence and achievement

Vocational preference has been frequently demonstrated to have little long-term reliability or realism in adolescence (e g , 613), although in college students it has generally proved more stable and realistic (775· 355) It is often asked, however, whether the scores of interest inventories provide one with information sufficiently different from expressions of vocational preference to justify the time and expense, and, as the preferences of some groups of college students have proved rather stable, it has been suggested that in their case the inventories may be of little value (926) Counselors working with clients in schools, colleges, and guidance centers frequently comment on the large number of cases in which

Strong's Blank merely confirms what one already knew from interviewing the client and, what is more, what the client already knew himself It is therefore pertinent to inquire concerning the relationship between Strong scores and expressed preferences some relationship would presumably be evidence of validity, while a nearly perfect relationship would suggest substituting a single question for the whole inventory

In two investigations (450,719), the conclusion was drawn that scores on Strong's Blank were less useful than expressions of preferences In both instances the conclusion was based on low correlations between inventory scores and vocational preferences of high school girls, and on the tendency of the former to be more concentrated in a few fields than the latter But both studies involved the Women's Blank, in which the clustering of scores has been seen to be due to the strength of one factor which is common in women and in many women's occupations

Bedell (59) found that only two of 17 women's scales had correlations of more than .50 with the self-estimated interests of freshmen women Data for 1000 men at the University of Minnesota were analyzed by Darley (189 21–25), with a resulting contingency coefficient of .43 between claimed vocational choices and inventoried interests as determined by his classification of Strong scores into primary, secondary, tertiary, and no interest patterns An examination of the basic data is perhaps more revealing of the inadequacy of expressed preferences as indices of measured interests Scientific choices were indicated by 374 men, of whom only 71 had primary measured interests in the scientific field, 214 had no primary interest patterns, 45 had business detail interests, and the rest were scattered among the other fields, 137 claimed linguistic choices, of whom 26 had measured primary interest patterns of that type, 65 had no primary patterns, 21 had social welfare interest patterns, and the rest were scattered, 169 claimed business detail preferences, while 60 had measured primary patterns of that type, 69 had no primary pattern, 16 had business contact patterns, and the rest were scattered throughout other categories Allowance must be made for the fact that many had secondary patterns in the field of their claimed interests, but even then the discrepancies are substantial Moffie (534) worked with NYA boys averaging 18 7 years of age, who rated their interests in the fields assayed by Strong's Blank and were scored with Strong's group and specific scales the correlations ranged from — 07 to .47 and from — 05 to .54, respectively Moffie's explanation is that lack of maturity and experience on the part of adolescents invalidates their judgments of their interest in

different types of work, while the pattern scores of an inventory succeed in tapping their interests more adequately It might be suggested, also, that this lack of experience and insight is greater in some areas than in others. Some occupational fields, e g teaching, are more open to observation by the average youth than others, making easier the formation of preferences on the basis of interest, while others such as certified public accountant (which has the least reliable scale so far developed) are not so readily observed Great variations in the agreement of ratings and inventory scores of individuals were found by Arsenian (30), further substantiating the hypothesis that maturity and experience, which vary from one person to another, account for the differences in agreement between measured interests and preference or choice Finally, data reported by Wrenn (942,943) show that the more intelligent college students are more likely to "choose" occupations in which they make high scores (45 percent of the superior group rate A on chosen occupation, 3 percent C), while the less able are more likely to make low scores in their preferred occupational field (22 percent rate A, 20 percent C) This suggests either that the more able students have more insight into their interests than the less able, or that their superior verbal ability enables them more adequately to integrate their rationalizations concerning interests Whether or not we are dealing with rationalizations or insights can be ascertained from the extent of the relationship between inventory scores and objective criteria such as completion of training, grades, and stability of employment in a field

The relative predictive value of inventoried interest and expressed preference has been studied only by Wightwick (926), who found that 44 percent of 115 college women were employed in the field of the freshman choice four years after graduation, and 73 percent in the field of their senior choice, in contrast with 58 percent employed in occupations in which they had as freshmen made A or B+ ratings This led the author to the conclusion that measured interests are not as valid predictors of vocational choice as expressed preferences, conclusion which seems rather odd as it can be based only on a comparison of *freshmen* inventory scores with *senior* preferences (58 vs 73), a comparison of freshmen test results and freshmen preferences suggests, instead, that inventories are superior to expressed preferences (58 vs 44) The greater validity of senior preferences is no doubt due to the nature of the criterion field entered It is to be expected that senior preferences would reflect an element of realism, including considerations of finances, opportunities, and family pressures

which would make them perhaps less valid indices of interest than test scores, but more valid predictors of occupation entered

Unfortunately Strong's nine- and ten-year follow-up studies (775 393–403) have not been analyzed in the same manner as Wightwick's They do show that about three-fifths of his college students were employed in the field of their freshman or senior choice five and ten years after graduation, they also show a substantial relationship between interest scores in college and field of subsequent employment, as seen in the discussion of the permanence of interests and as brought out below in the material on job satisfaction, but the data are not so organized as to show what percentage of men entered and remained in fields in which they made A, B+, or lower scores

The relatively low correlation between expressed preferences and inventoried interests in high school, the tendency of the less able students to prefer fields in which they lack measured interests, and the superiority of inventories to the expressed preferences of college freshmen in the one known study which has made such a comparison with objective evidence as a criterion, suggest that inventories can improve the quality of counseling and prediction With college upperclassmen and adults expressed and inventoried interests will probably generally be found to agree, but in some cases insights of this type are lacking, especially when external pressures have been at work on the client

Vocational achievement has served as a criterion of the validity of Strong's Blank particularly in work with life insurance salesmen It might be argued that the criterion of the validity of an interest inventory should be satisfaction, rather than achievement, certainly satisfaction should be one of the outcomes of interest But if interest produces satisfaction it should also result in achievement, granted the necessary abilities, for the satisfied worker should throw himself more wholeheartedly into his work This might not be true of all occupations, for theoretically there might be some fields in which the work can be done equally well regardless of interest and satisfaction in the work, provided the end-result (pay, prestige, etc) is desired, but in other fields the congeniality of the activities engaged in might be important to success

That insurance sales is one of these latter is indicated by a number of studies by Strong (772), Bills (90,91,92), Ghiselli (287) and others, most of which are summarized in Strong's book (775 487–500) Only illustrative data are therefore considered here

Only one of Strong's studies used as subjects a group of applicants for

employment as insurance salesmen, the other groups consisting of men already employed or, in one instance, released, by their company (775 487–488) In the pre-employment study, the applicants were tested in a small agency, 20 were employed, and only 16 remained more than three and one-half months The data of the pretested group are therefore not very conclusive, although they do show a clear tendency ($r = .48$) for the higher-scoring men to sell more insurance When data from all groups, all agencies, were combined the relationship between interest scores and sales (criterion reliability = .81) is as shown in Table 30, adapted from Strong (775)

TABLE 30

PERCENTAGE OF AGENTS IN EACH LIFE INSURANCE INTEREST RATING WHO PRODUCE $0 TO $400,000-AND-UP ANNUALLY

Annual Production	*N*	*Percent in Each Rating Producing*				
		C	*B−*	*B*	*B+*	*A*
$0 to $ 49,000	38	52	33	27	22	9
$50,000 to $ 99,000	52	24	17	45	34	16
$100,000 to $149,000	31	18	17	14	7	19
$150,000 to $199,000	37	6	17	9	13	22
$200,000 to $399,000	47	0	17	0	20	31
$400,000 up	6	0	0	5	4	3
Total		100	101	100	100	100
Number	211	17	6	22	45	121

There is a rather clear tendency for those who made high scores to be those who sold the most insurance 56 percent of the A men sold enough insurance to make a living by then-current standards ($150,000), as compared with only 6 percent of the C men Although the coefficient of correlation for 181 of these cases is only .37, the relationship is statistically and psychologically significant, for it must be remembered that most of the men were tested after a long period of employment, after the low-producing and low-scoring men had been eliminated by natural selection The greater range of scores and sales which would characterize applicants would undoubtedly yield a higher correlation coefficient If these men had all been tested as applicants for employment, or better still as college students, and had made similar scores, findings would be quite convincing As all but 16 of them were tested after they had been on the job some time, it is possible, however, that some of the poorer salesmen indicated liking for fewer of the sales items than did their more successful fellows, not because they actually liked sales work less, but because they were somewhat dissatisfied with the financial results of their work Apparently Strong has not taken this possible rationalization of

failure into account, for he makes no mention of the possible differences between pretested and posttested responses But even if such forces were at work, the relationship between inventoried interests and success in selling life insurance is noteworthy

The life insurance and real estate salesmen's scales of the Vocational Interest Blank were combined in Bills' study of 588 newly employed casualty insurance salesmen and compared with ratings of success after one year on the job She found that 76 percent of those who made low scores were failures, while only 22 percent of the high scoring group failed Ghiselli worked with a much smaller group of casualty insurance salesmen, 29 in all, finding significant relationships for the CPA and occupational level scales (38 and 27) He reports that they tended to make high scores on the business contact and detail keys, but that contrary to Bills' findings the contact scales did not correlate with performance As his cases are far fewer in number, the relation can hardly be considered disproved by this one study

Another type of salesman, selling detergents on a wholesale basis over large territories and acting as service men on related matters (service time correlated 51 with profits), was investigated by Otis (580) The group was necessarily small, as there were few territories and the turnover rate was low (N = 17) His criterion was selling cost, with which the combined life insurance and real estate salesmen's scales correlated -50 With numbers as small as these the data are merely suggestive, but promising

Accounting-machine salesmen, 143 in number, and 283 service men of the same types of machines, were studied by Ryan and Johnson (660) They found that the two groups were differentiated from the general population by especially constructed standard-type scales, but that scores on these scales had no relationship to success They then developed another set of scales based on the differentiation of successful from unsuccessful men in the occupations in question These scales did differentiate other groups of successful and unsuccessful men in the same jobs, the critical ratio for the service men being 4 8.

The relationships between interests and achievement in several other occupations have been summarized by Strong, often from unpublished studies, it is from him (775 501–504) that the following are taken, except when otherwise indicated.

Psychologists who were starred in *American Men of Science* averaged 48 7 on the psychologist key Strong explains this slightly below average

score on the basis of the low scores made by some applied psychologists, two of whom scored below 30 and later went into business, but this explanation seems unnecessary in view of the expected shrinkage in the means of new groups when criterion and norm groups are one and the same It can therefore only be said that eminent psychologists taken as a group do not seem to differ from somewhat less eminent psychologists (the Fellows of the standardization group)

Teachers were rated by Ullman and by Phillips for success of performance, the ratings did not correlate with interest inventory scores

Engineers rated as outstanding by an engineering dean were compared with full and associate members of the four engineering societies The outstanding engineers made higher scores than the associates

Aviators who failed in flying training were not significantly lower on the aviator scale than were those who were successful in training, perhaps because of the small size of the samples Another set of pilot scales were constructed in a study initiated by the writer in the Air Force (316: 608–611) A total of 650 aviation cadets were tested with the Vocational Interest Inventory, and scales were developed on the basis of item validities, the scale based on even-numbered cases being cross-validated on odd-numbered cases, and vice versa The correlations between these scales and success in primary flying training were insignificant (— 03 and — 10), confirming what Strong found with smaller groups

Advertising men, 36 in all, were rated by three officials of their agency Although the significance of the relationship was not tested, the men with higher ratings tended to have higher scores on Strong's advertising scale

Foremen, 59 of those employed by a large chemical plant, were rated for characteristics which are not described The correlations between ratings and Strong scores were 34 for chemist, 31 for engineer, 25 for CPA, and — 31 for life insurance salesman These relationships are such as might be expected in a sub-professional technical job, except that with CPA Thirty others were tested by Schultz and Barnabas (682), and were rated for budget-control efficiency and employee relations The correlations between Strong's scales for production manager and occupational level, on the one hand, and combined ratings were respectively 38 and 22

Janitor-engineers rated above average in their work ($N = 44$) were found by Berman, Darley, and Paterson to make higher scores on the technical and scientific but not on other scales than did a group of 23

who were rated below average In the same study 123 policemen rated by their captain were found to be differentiated on the basis of scales which measure interest in social contacts

Summarizing the evidence on the relationship between inventoried interests and success in an occupation, we have seen that it is significant in the case of several quite different types of sales jobs, although in some this is so only when success-failure rather than occupational-differences keys are used Success in psychology and in teaching were not related to the degree of similarity of interests to those of persons employed in those fields, but success in advertising, technical foremanship, janitorial work, and police were Successful and unsuccessful aviators were not differentiated by success-failure scales

The sales data are consistent with the writer's hypothesis concerning interest and achievement, for selling life insurance requires a substantial degree of self-direction and willingness to persist in the face of a cool welcome, presumably only a person who finds a real challenge in locating prospects and in making himself pleasant and helpful to them could make enough calls to earn a living Congeniality of the work is important, and there is a significant relationship between interest and achievement The same is true of casualty insurance salesmen, and of wholesale salesmen in whose work service to customers is an important function But in somewhat more routine sales work interest is related to success only when the interests of successful men in the occupation are contrasted with those of failures in the same field rather than with those of men-in-general

The other findings are more difficult to synthesize or rationalize The apparent contradictions may lie in the differences in the criteria of success being starred in *American Men of Science* for one's research contributions is not comparable to being rated highly for the successful management of advertising accounts Perhaps advertising is partly a sales occupation (r life insurance salesman = 59), in which case the importance of interest is explainable in the same terms Psychology and teaching are non-competitive, and success in both fields can be achieved in a great variety of ways, perhaps congeniality is less crucial to them because of their varied outlets Just why interest should so clearly play a part in success in non-competitive fields such as foremanship, janitorial work, and police is difficult to see Congeniality could be important, in that all three groups have to put up with the vagaries of a variety of people, but then so do teachers More studies, with more detailed analyses

of the duties of those involved, are needed before the significance of these findings will be clear

Occupational differentiation being the basis on which the Strong Vocational Interest Blank was constructed, most of what might be covered in this section has already been dealt with in earlier sections, particularly those concerned with the construction of the occupational interest scales Some occupations have been studied without scales having been developed for them for example, Bluett (102) ascertained the patterns of interest scores characterizing vocational rehabilitation officers The applicability of the adult scales to pre-occupational groups was verified by Goodman (296), who found that engineering students differed in the expected ways from liberal arts students, and by Barrett (45), who found that women college students majoring in art made higher scores on the artist scale than did other students But the most significant problem still to be discussed is that of the differentiation of women's vocational groups on the basis of their interests

Women's and girls' interests have been investigated with Strong's Blank by Laleger (450), Skodak and Crissey (719), Crissy and Daniel (182), and others besides Strong himself (775 162–168) These studies have shown that it is more difficult to differentiate women on the basis of their interests than it is men The manual for the Women's Blank shows a surprisingly large number of substantial correlations between occupations which would not, on the basis of data for the men's form, be expected The correlation between the women's office worker and nurse scales, for example, is 55, while that between office worker and housewife is 84 It has frequently been noted that populations of high school girls and college girls tend to make far more high scores as nurse, office worker, elementary school teacher, and housewife than should be found in a random sample Stuit (783) found this even among teachers' college students A suggestion as to why this may be the case emerges when it is noted that the correlations between the housewife scale, on the one hand, and those for nurse, physical education teacher, elementary school teacher, office worker, and stenographer on the other are respectively 59, 56, 84, 77, 80

The factor analysis by Crissy and Daniel (182) referred to earlier carries this thought further They found four factors in women's vocational interests, three of which were like those found by other psychologists in studying men, but one of which they called "male association," thereby bringing down on their heads a storm of protest from women

psychologists It is this factor which others have called interest in multiplicity of detail, interest in the convenience of others, interest in order, and non-professional interests It has a very slight loading in the masculinity-femininity scale Whatever the factor is, it seems to be present in a great many women, especially in those in the occupations named, and it is present in negative form in other women, particularly those who make high scores as authors, librarians, artists, physicians, and social workers It is worthy of note that the occupations in which the so-called male association factor is important in a positive way are those which may be entered after a relatively brief and easily obtainable education, whereas those in which it is of negative importance are by and large those which require a longer and less easily obtained education or which are entered only by the persistent and highly motivated It would be helpful to have the marriage rates in each of these occupations, in order to as certain whether or not those who are characterized by a strong "male association" factor do in fact marry in greater numbers Observation suggests that the loss of women office workers through marriage is greater than the loss of women authors, physicians, and social workers for the same reason, but this is no doubt partly because the latter groups frequently continue their work even after marriage If both groups of occupations marry with more or less equal frequency this factor can hardly be named "male association", and if it really is that, why is there no evidence of a "female association" factor in men, most of whom also marry? As the factor has been isolated only in women, is positively related to stopgap and negatively related to career occupations, and is more important in the occupation of housewife than in any other (factor loading 83), it is suggested that this is in reality a home-vs.-career factor. The home or career decision is one which many women have to make, and which most decide in favor of the home It is presumably the presence of this factor which makes it difficult to measure the vocational interests of women with Strong's technique, for it outweighs vocational interests in many instances As will be seen in connection with the Kuder Preference Record this difficulty is not an insurmountable obstacle to the measurement of women's interests, but to overcome it it is necessary to use a different type of inventorying device

Satisfaction in one's work has seemed to most psychologists and counselors to be the objective of counseling or employment on the basis of interest inventories. But the appraisal of vocational satisfaction is not a simple matter, for a multiplicity of factors are involved and not all

of them are easily accessible The criteria of vocational satisfaction in studies of interests have consisted of stability in the occupation (in contrast to the position), and expressions of satisfaction or dissatisfaction by the worker

Occupational stability was the criterion favored by Strong (775 §84–§88) and used in his follow-up studies It is reasoned that *interest* determines the *direction of effort*, *ability* the *level of achievement* The criterion of a vocational interest inventory should therefore be the extent to which it predicts the direction of effort College students who enter an occupation and remain in it for ten years after graduating from college are presumed to be interested in and satisfied with the direction of their efforts, even though a few are known to persist because of family or economic reasons Those who change from one field to another are presumed to do so because they find the first field of activity unsatisfactory, and expect that the second will prove more so, despite the fact that some individuals change fields of work for economic reasons If these assumptions can be granted, and they probably can in higher level economic groups such as graduates of a private university, then occupational stability is a good index of vocational satisfaction and a suitable criterion of the validity of a vocational interest inventory

The ten year follow-up (775 §93) consisted of 287 Stanford University seniors tested in 1927 and followed up in 1928, of whom 223 were retested in 1932, and 197 again retested in 1937 The nine-year follow-up was based on 306 Stanford freshmen tested in 1930, of whom 174 were retested in 1939 The principal findings and conclusions are as follows

1 *Men continuing in an occupation for 5 or 10 years after college made higher scores in it than in other occupations* (mean standard score 50 2 vs 47 7),

2 They tended to make higher scores in that occupation than did other men (data too complex to reproduce here),

3 They made higher scores in that occupation than did men who changed from that occupation to some other (standard score 48 0 vs 44 0),

4 Men changing from one occupation to another after employment in the first field did not make higher scores on the latter occupation when in college, but their average scores were substantially lower in both the first and the second occupation than were those of men in groups 1, 2, and 3, above (standard scores 42 4 and 40 5), which suggests that those who change occupations have less clearly defined

interests, or less insight into them, than do those who remain in the occupation of their first post-college choice

The hypothesis that interest inventory scores are manifestations of stereotypes does not seem to be sufficient to explain away these findings It could, if true, remove the significance of the first finding, since an unchanging stereotype would be the result of staying in the same occupation, it could do the same for the second finding, for the men who enter a given occupation would be expected to have the relevant stereotype to a higher degree than others, and it could be argued that the men in group three who changed to other fields did so because they found that their concept of the occupation and of their role in it did not coincide with the facts, but the fourth finding and conclusion imply that an interest pattern is the result of more than a mere stereotype or even a more deep seated self-concept, but rather the product of a more fundamental combination of personality traits, aptitudes, and modifying experiences This fourth group apparently lacked the highly organized personalities (in the broadest sense) which characterized the other groups, as indicated by the mean standard scores, even after several years of occupational experience in which they might have acquired the stereotype The lack must have been one of aptitudes, temperament, and values

Women were followed up eight years after testing and four years after graduation from college by Wightwick (926) Of her 115 subjects, 58 percent were employed in occupations in which they had made A or B+ ratings, while 77 percent were in fields in which they had at least tertiary patterns The data were not analyzed for stability of employment in the same way as in Strong's study, but in 1941, 43 percent were employed in occupations in which they had made A or B+ scores in 1933

These findings seem to be confirmed by trends brought out in a study of 76 adult men by Sarbin and Anderson (667), in which the client's statement of vocational satisfaction or dissatisfaction was related to his primary interest pattern They found that 82 percent of the men who expressed dissatisfaction with their current occupations did not have primary interest patterns in the fields in which they were employed, but there is no indication as to how many satisfied workers possessed primary interest patterns in the field of their endeavor If their data are recomputed to permit another comparison, it appears that 57 percent of those who had a primary interest pattern in the field of employment were dissatisfied, as compared with 52 percent of those who lacked the appro-

priate primary interest pattern This would seem a strange finding, were it not that the subjects were clients of an adult guidance center, and therefore were, as might be expected, a predominantly dissatisfied group Although Sarbin and Anderson's statement that "adults who complain of occupational dissatisfaction show, in general, measured interest patterns which are not congruent with their present or modal occupation" (667 35) is exactly what one would expect to find, it can hardly be said that they have demonstrated the truth of the statement

Satisfaction in a professional curriculum was correlated with inventoried interest in that field by Berdie (77), in a study of 154 engineering sophomores who had been tested as freshmen Satisfaction was measured by a modification of Hoppock's Job Satisfaction Blank, in which the term "curriculum" was substituted for "job" and "occupation" The correlation between scores on Strong's engineer scale and satisfaction score was 10, too low to be significant When the data for 43 men whose blanks had been scored for all occupations were subjected to analysis of variance, it was found that those with no interest pattern in the engineering field were significantly less satisfied than those with a primary, secondary, or tertiary pattern in the physical sciences The numbers were so small, however, as to make conclusions highly tentative

Although the evidence concerning interest and job satisfaction which consists of occupational stability data is impressive, there is a need for further studies using clinical and psychometric indices of vocational satisfaction Sarbin and Anderson's study was a step in this direction, as was Wightwick's, but an adequate investigation of clinically or psychometrically determined vocational satisfaction in relationship to inventoried interests has yet to be made

Use of Strong's Vocational Interest Blank in Counseling and Selection The findings of research which have been reviewed in the preceding sections have shown that interest is not a completely independent entity, but rather something which is related to general ability, special aptitudes, and values in various ways Linguistic and scientific interests are positively correlated with intelligence, technical interests are related to mechanical aptitude, and business interests are related to the tendency to stress material as opposed to theoretical, social, or aesthetic values, to cite just a few of these relationships But the very complexity of these relationships supports the hypothesis that interests are sufficiently unique to warrant special consideration in the study of an individual or a group, and other evidence shows that they have significance in and of themselves which makes their study important It seems likely that aptitudes,

values, and perhaps temperament are fundamental factors which, together with experiences in childhood, determine the development and nature of interests, but the end result is a type of individual differences which take on a character of their own There seems to be something magnetic about interests, pulling people in their direction and holding them in place once there

The development of interests has been seen to be well under way by adolescence, for by age 14 or 15 the interest patterns of boys and girls have begun to take forms similar to those of adults, and these patterns are generally modified by increasing maturity by becoming more clear cut, and by a tendency, in boys at least, toward great socialization of interests By the time boys and girls are from 18 to 20 years of age their interests are fairly well crystallized, and in most cases change very little thereafter

The occupations for which Strong's inventory has been validated are primarily professional, managerial, and clerical, although a few skilled occupations are included among the scales Its usefulness is therefore primarily with those persons whose intellectual and educational level is high enough to provide a sound basis for aspiration to the middle or upper half of the occupational ladder The men's form can be scored for about 40 occupations, the women's for more than 20, while these seem like very few, compared to the large number of jobs which have been differentiated in other ways, the limitations of the instruments are not as great as this suggests The occupations are more broadly defined than in the *Dictionary of Occupational Titles* (888), for example, and what is more, the intercorrelations have shown that they fall into interest families, that these occupations can be grouped according to common underlying interests This means that by using this inventory and scoring it for a relatively small number of occupations one can tap interest in a few core fields in which most known occupations could probably be placed It is important to bear in mind, however, that interest is not necessarily a predictor of success, even when needed abilities are present, for interest seems to be related to success only when the congeniality of the activities in question affects application, and when the effects of application are readily determined, as in competitive work such as sales It seems to be much more likely to be important to satisfaction and stability in a field than to quantitatively judged success The women's form is not as satisfactory as the men's, because of the commonness of one interest factor in women, it is only in the cases of those with clear-cut career interests that it is likely to prove valuable

In school and college the Vocational Interest Blank is sufficiently well

understood by 10th graders to be used with them, and their maturity is great enough to give their scores meaning despite the fact that there is some subsequent modification of interests The occupational interest scores have value for the prediction of educational achievement when screening on an ability basis has already taken place, and when there has been no screening on the basis of interest In most situations, however, the choice of curricula or courses by students gives them enough of an elective character to nullify the relationship between interests and grades Completion of a sequence of courses or of professional training is, however, related to interests as measured by the Strong Blank, for those whose interests are unlike those of people in the same occupational field tend to drop out more frequently than do students with appropriate interests The inventoried interests of high school students are of more value in vocational diagnosis than are their expressed preferences, on the other hand, the preferences of college students are likely to be mature enough to warrant more serious consideration, and are likely to be not much less significant in freshmen, and slightly more significant in seniors, than measured interests The younger and less able the boy or girl, the more need there is for a good interest inventory, Strong's Blank seems to meet this need within the normal ranges of male high school juniors and seniors, college students, and adults, it does so less well for girls and women, because of the career-vs-home factor The older and brighter the individual, the less likelihood there is that Strong's Blank will reveal anything new to the subject, although the confirmation of interests is often very helpful and new light is sometimes thrown on confused or poorly understood situations

The counseling use of Strong's Blank in school and college can therefore be both for choice of curriculum and for choice of occupational field Students may be encouraged to major in fields in which they have primary interest patterns, with the knowledge that they are more likely to complete work in those fields than in those in which their interests are not so strong Their choice of occupations for which they have appropriate measured interests may be viewed with more confidence that they will still prefer those fields after five or ten years of employment in them Despite the possibility of faking scores in an attempt to impress, the inventory has similar value in student selection programs

In working with high school and college students one not infrequently encounters cases in which there seems to be no primary interest pattern. As these are usually students or clients who have no clearly defined ex-

pressed preferences and who hope that the interest inventory will discover some hidden interest, this experience is one which is especially frustrating to novice counselors The frequency of such cases in a college population has been investigated by Darley (189 19–21, Ch 5), who found that slightly more than 52 percent of 1000 University of Minnesota students had no primary (A and B+) interest patterns, while 16 percent had only tertiary (B and B−), and 3 percent had no distinguishable, interest patterns Darley set up the hypothesis that students with high interest maturity and no primary interest pattern would make poorer grades in college than students who had primary interest patterns, but this was not verified by his evidence As he puts it, "the case with no primary pattern will continue to be clinically difficult for the counselor as usual, more and better research is necessary " Strong showed that the interests of business students are less clear cut than those of professional students (775 420), he suggests that people with widespread interests and often without primary interests should consider business, particularly if they have secondary interests in the business groups (775 430), but he, like Darley, ends by giving up "These are the hardest of all people to counsel, because they have so little to contribute and either they have a lot of half-baked plans that change from interview to interview or they sit back and expect the counselor to prescribe the remedy" (775 441) In the writer's experience with college students it also seemed that the undifferentiated students were those who entered business, for lack of something more challenging He is somewhat reluctant to let the matter rest there, however, in view of Strong's findings concerning the differentiation of people at lower occupational levels when a different point of reference is used Research in the "undifferentiated" group both in college and elsewhere should presumably be pressed, using other points of reference than that of the standard scales

In guidance centers the counseling use of Strong's inventory is similar to that in schools, with the exception that there it is often given to entire classes as a part of a routine testing program, whereas in a guidance center it is part of a tailor-made battery individually administered In mass testing which has been properly motivated the examinee's answers are likely to be frank and free, for even though motivated to co-operate he is likely to feel that he has relatively little at stake In the individual testing program there is more liklihood of self-scrutiny and of unconscious warping of responses to make them congruent with an acceptable self-concept In the former case scores may not be as high, but they reveal

the patterning of specific interests more truly, in the latter case, they are more indicative of self-concepts Both types of data have their value, provided the counselor knows with what it is he may be working

A guidance center has an advantage over employment services and departments in the use of inventories such as this, in that its functions are recognized as being more advisory than administrative, the former role being one which encourages frankness on the part of examinees Despite this fact, consultants making evaluations need to be alert for case history material which tends to support or to contradict the evidence of the inventory It might be well if two inventories, known to differ in their transparency, were used, to provide an index of tendency and of direction of distortion of interest scores The research necessary to the development of such an index has not been carried out as yet, but the germ of the idea is to be found in a paper by Paterson (586)

In employment services inventories such as this are rarely used, as the type of counseling offered there has generally to do with employment rather than with choice of a field of work, and the interests of employment applicants have generally seemed assessable by less complex methods As more attention is paid to the needs of inexperienced youth, on the one hand, and to the careful appraisal of adults applying for competitive jobs, on the other, interest inventories should probably find more use in employment services

In business and industry the use of Strong's Blank has been confined to the selection testing of applicants for sales positions, particularly those in which the importance of the congeniality of the work, the independence of the salesmen, the intangibility of the item sold, or the competitive nature of the selling have been notable These items include life insurance, casualty insurance, real estate, business machines, and vacuum cleaners As work with this type of instrument began in an attempt to distinguish sales engineers from technical engineers one would expect to see other successful applications made as time goes on Here, more even than in guidance centers, the possibility of faking unduly high scores needs to be considered The indications are that despite this tendency the Blank is a useful sales selection instrument, an index of distortion such as that suggested above would make it even more so

CHAPTER XVIII

OTHER MEASURES OF INTERESTS

The Kuder Preference Record (Science Research Associates, 1939, 1943, and Short Industrial Form 1948)

WORK with this inventory was initiated by Kuder at Ohio State University early in the 1930's, leading to the publication of the inventory in 1939 Three forms were tried out during this experimental period After the 1939 edition had been in use for several years it seemed desirable to cover mechanical and clerical activities more adequately, and the second edition was developed and published, incorporating also a change in the form of the items A short form for use in business and industry was published in 1948 Publication of the inventory was welcomed by many counselors in schools, colleges, and guidance centers, because it was more economical to score than Strong's Vocational Interest Blank, then practically the only inventory which had been well validated, and because it also showed signs of having been subjected to a good deal of research. Furthermore, its format and marking device had an immediate appeal to students taking it Users of vocational tests therefore often included the Preference Record in their batteries, interpreting its results in very much the same terms as those of Strong's Blank, simply on the basis of the general similarity of the types of items and scores, which seemed like those of Strong's group scales Today the Kuder is one of the most widely used vocational tests and inventories, and additional evidence concerning the nature and vocational significance of the traits it measures is published practically every month in the professional journals

Applicability The Kuder Record was designed for use with high school and college students, and with adult men and women The items were so written as to be applicable to both sexes, the vocabulary was kept as nearly as possible at the high school level, and the content seems to have been selected for its familiarity to adolescents as well as to adults Two reports on the suitability of the inventory for high school students have been published Christensen (157) tried it out on 27 9th graders and

ascertained that many of the items were not understood, when the class was instructed in the meaning of the items and retested, the scores changed appreciably The reading difficulty of the Kuder was checked by Steffire (752), who used the Lewerenz formula for vocabulary grade placement, he found that the vocabulary difficulty grade level was 8 4, and that it is easier than that of the Strong (10 4), the Allport-Vernon Study of Values (11 3), and the Cleeton (12.0), but somewhat more difficult than that of the Lee Thorpe (6 8) and Brainard (6 4) These findings suggest that the Kuder can be administered to typical 8th grade boys and girls, although the less able will have difficulty with some items its use at the 9th of 10th grade levels is likely to prove satisfactory in this respect Norms are available for the interpretation of the inventory with high school students and adults

The transparency of the items in the Kuder, or the ease with which faking and unconscious distortion of responses can take place, has seemed a problem to many users That their objections have some basis in fact is suggested by the nature of the items, as inspection reveals them to have rather obvious vocational implications Both the Kuder and the Strong inventories were administered to a clerical employee being considered for transfer and promotion to a desired personnel position by Paterson (586), who compared the man's responses on both forms The data suggested that the employee's interests were truly clerical, that he wanted to appear in the best possible light as a potential personnel appointee, that his scores were distorted in the direction of personnel interests by this fact, and that the Kuder was more affected by distortion than the Strong As these are merely observations of one case they are not conclusive, but they do seem to confirm the general opinion of users of vocational interest inventories Two experiments designed to test the transparency of the two measures, as Strong tested that of his own, have been completed Bordin (112) has reported one such, in which it was found that the professed social service and literary interests of college students were more highly correlated with Kuder than with Strong scores (e g , $r = 43$ vs $r = 29$), suggesting greater transparency in the Kuder, but the trend was not consistent for other scales Cross, in an unpublished study of high-scoring students (181 males, 183 females), found clear-cut evidence of ability to lower and to raise Kuder scores according to directions

Another aspect of this question of the meaning of inventory items and the orientation of respondents was investigated by Piotrowski (608), who tested 18 superior students in a school of social work with the Kuder and

the Rorschach All subjects scored high on the social service scale, but psychiatric interviews led to the conclusion that only 11 of the inventory scores were "valid," while 7 were "invalid", in other words, 7 of the social workers were not genuinely interested in social welfare, but made high scores because of conscious or unconscious distortion The Rorschach responses of the two groups were then compared, with the conclusion that those who really had social service interests (as confirmed by interview data) were closer to reality, had a wider range of psychological experiences, were more realistic in their aspirations, more interested in people for their own sakes, more self-confident, and less frequently subject to despondent moods While the results for other preoccupational or occupational groups might reveal fewer invalid scores (assuming the validity of the psychiatric interview) than a field such as social work, the evidence does indicate that distortion of scores on the Kuder can seriously affect the results

There is, finally, the question of changes in responses to this type of inventory with increasing age Although one would expect Strong's findings to hold for interests however measured, there is the possibility that the form of the question and the method of scoring affect findings in the case of a particular instrument, making such generalizations unsafe until appropriate evidence is adduced Retest reliabilities after a lapse of 15 months were computed for 16 adult subjects (ages unreported) by Traxler and McCall (868), who found that they ranged from 61 for social service interests to 93 for musical interests, the median being 83 This suggests a considerable degree of stability of responses DiMichael and Dabelstein, in an unpublished paper (200) found reliabilities ranging from 70 to 89 Even for the least reliable scale letter ratings (A = 75 percentile or above) changed in only 9 percent of the cases Traxler and McCall, and Kuder in his manual, provide data showing that the changes which take place during senior high school and college years are relatively slight, making unnecessary the use of special norms for each high school grade This conclusion cannot be compared precisely with Strong's, as his norming procedure was different, but it does appear to be at variance with it Strong's and Carter's work, reviewed elsewhere, showed more convincingly that certain changes do take place in the interests of adolescents and that they are fairly well crystallized by the *end,* rather than by the *beginning,* of the high school years, they have merely begun to take shape by age 14 or 15 Until intensive work on age changes has been carried out with the Kuder, it seems wise to assume that some changes such as those

known to take place in responses to Strong's Blank also affect Kuder scores

Content The Preference Record consists of preference items arranged in triads Item four illustrates the principle

Build bird houses
Write articles about birds
Draw sketches of birds

The examinee decides which of these three activities he likes best and marks it to show his first choice, then he decides which he likes least, and marks it to show his third choice The activities in each item are so written as to tap three or more different types of interest, in this case *mechanical, literary,* and *artistic* There are 504 such items (standard form), assessing interest in a total of nine[1] different types of interests

Administration and Scoring There is no time limit, as there are no right or wrong answers, the time required by high school students is from thirty minutes to one hour, by college students approximately forty minutes It is necessary to make sure that the directions for using the response pins are correctly followed, but, as examinees are usually intrigued by the mechanics of the inventory, motivating them to follow directions is relatively easy Scoring may be done by hand, using appropriate answer sheets and a pin to prick answers, in which case the common procedure is to have examinees do the scoring themselves The directions are clear, and it takes about fifteen minutes to obtain all nine scores Profile sheets are provided on which examinees convert their scores to percentiles and plot them graphically This method has generally been found to be a good device for getting pupils interested in their scores and to provide a springboard for discussion of vocational interests Machine-scoring is also possible, with the use of special answer sheets The scores obtained are for mechanical, computational, scientific, persuasive, artistic, literary, musical, social service, and clerical interests

Norms The 1946 edition of the manual contains norms for three different base groups. The first consists of approximately 2000 boys and 2000 girls, in grades 10, 11, and 12, the three grades being lumped together because of the lack of important grade differences but the sexes separated because sex differences are significant The second is made up of adults engaged in a variety of occupations 2667 men from 44 occupations, and 1429 women in 29 occupations, again treated separately because of sex differences Thirdly, there are norms for college students, those

[1] A tenth, 'Outdoor,' interest scale has been added

for women based on 1263 students in various curricula, while those for men are, for the time being, derived from groups of about 200 each from several different colleges The profile sheets provided with the answer sheets are based on the first two groups, and Kuder expects to provide one for the third group While these norm groups are helpful in providing a backdrop against which to view the interests of an individual, their composition is not as vital a question as in the case of Strong's Blank, for with the Kuder one studies the relative strength of each of nine different interests *within an individual,* whereas in the Strong the comparisons are basically *between groups of individuals* classified by occupations

Having obtained a profile of scores which shows the relative strength of the different types of interests in the person being examined, the next question which arises is that of the occupational significance of the profile It was the absence of occupational norms which made many users of vocational tests hesitate to use Kuder's inventory, despite the care with which it was constructed and the economy with which it could be used It was not until after World War II, for example, that the writer used it in counseling on anything other than an experimental basis, just because it did not seem sufficient to know that a client was more interested in mechanical activities than in any other type, when what counts in vocational adjustment is how his interests compare with those of persons who have succeeded in the field This point has effectively been made by Diamond (199a), in an important study of the occupational significance of Kuder percentile scores

The 1946 manual has to some extent made good this deficiency by providing norms for 44 men's occupations and 29 women's, supplemented by curricular norms for women college students in 24 different fields The numbers in any one group are small, ranging from 16 men English teachers and 16 women language teachers to 185 male meteorologists. Strong's work suggests rather clearly that these numbers are too small to be reliable, but they are better than no data at all, and an unpublished study by Triggs has shown that for one group, at least, the adding of additional cases makes no difference She tested 826 nurses, and found that their mean and sigma differed little from those for Kuder's group of 185 As the manual indicates that the test-author is interested in receiving additional occupational data for norming purposes, it may be assumed that better occupational norms will become available in due course Judging by the incomplete evidence in the manual, there is one other possible defect in the occupational norms that of sampling This problem has

been amply discussed in connection with Strong's work, so it need only be pointed out here that better evidence needs to be supplied concerning the type of employment, skill levels, degree of permanence, level of attainment, and regional location of the representatives of any given occupation The manual does not mention these variables

The occupational norms consist of the means and standard deviations of each occupational group on each interest scale, and graphic profiles based on these same means The profiles permit a more rapid inspection of the data than do means, and enable the counselor to compare quickly his client's profile with that of the various occupational groups As the work of the Minnesota Employment Stabilization Research Institute (223) and the United States Employment Service (225) has demonstrated, however, this technique has serious defects Not only is it impressionistic rather than exact, but the criterion upon which judgment is based is unsound, for the counselee is compared with the average person in the occupation rather than with the marginal worker To put it concretely, if the counselee is significantly below the mean of the occupational group at two points of the profile, and significantly higher at two other points, does that mean that the choice of that field would be unwise? It would be more helpful to know the critical scores for each trait being measured, for then a "low" score would be known to indicate a critical lack of a trait which has been found to be related to success or satisfaction in the occupation in question This is the procedure now used by the United States Employment Service in its General Aptitude Test Battery (225, see also pp 358 ff) Diamond's data (199a) are again highly relevant

To provide a less impressionistic method of comparing individual profiles with those of persons established in various occupations, Kuder has developed *occupational indices* which are a statistical summation of the similarity of the examinee's interest profile to that of the occupation in question The principle is similar to that used by Strong, although Strong applied it to a series of items whereas Kuder applied it to scores on a series of scales Only one occupational index has so far been published, that for accountant auditor (446) Triggs has also developed indices for nurses in several specialities, described in an unpublished paper As more of these indices are published the value of the Kuder in vocational counseling will increase, but the counselor will be called upon to exercise a high degree of judgment in deciding when a deviation from the mean is so great as to suggest the abandonment of an objective by the client

Standardization and Initial Validation Many users of the Kuder Pref-

erence Record have been puzzled by the method of weighting the items in the inventory The mental set established by Strong's work led them to believe that Kuder's interest scales were occupational in nature, that his scientific scale, for example, was the scale for a scientific family of occupations But the early manuals and published studies showed no evidence of occupational standardization The alternative explanation seemed to be that the keys were based on a factorial analysis of interests, such as were made with Strong's data, but again there was no evidence of such work Lacking any such empirical basis, the scales were not infrequently suspected of being the product of nothing more than *a priori* reasoning

Succeeding editions of the manual have attempted to make clear exactly how the scales were developed, but the writer has talked with competent applied psychologists who had still not grasped the procedure, simple though it is The first step was the construction of *a priori* scoring keys, in one of which all seemingly literary items were scored, in another all scientific, and so on The second step was to score the blanks of several hundred persons with these scales The third step was to make an item analysis, to ascertain the internal consistency of these scales If it was found that those persons who had made high *a priori* literary scores tended to choose a given item more often that those who had made low scores, the item was retained in the literary scale, if it was not so chosen, it was discarded After this procedure had been applied to all the *a priori* keys it was found that some of the empirically purified scales (the seven published with the first edition) were internally consistent, independent of each other, and reliable, while others (atheletic, religious, and social-prestige interests) were not internally consistent or independent—they were, in fact, purified out of existence by the item analysis (the social-prestige scale actually split in two) The item analysis therefore gave an empirical basis for stating that the interest scales measure something, and that these entities are independent of each other and unchanging in their composition The method of naming traits is then comparable to that in factor analysis, and depends on inspection of the items and judgment as to their nature The names given by Kuder seem warranted, as might be anticipated when the items are rather transparent The two scales added with the second edition were for mechanical and clerical interests, and were based only on internal consistency, they are correlated somewhat more highly with the scientific and computational scales

The intercorrelations of the original seven scales range from −.34 for

the scientific and persuasive scales to 19 for the scientific and computational scales, when based upon 2267 adult men in a variety of occupations (446) The somewhat higher intercorrelations for the new scales are 50 for the clerical and computational, and 405 for the mechanical and scientific scales

Reliability The reliability of the Kuder scales has been ascertained for several different age groups and summarized in the manual by Kuder For 8th-grade students the Kuder-Richardson reliability coefficients range from 84 to 96 (100 boys and girls), for 125 high-school senior boys they range from 87 to 93, for a similar number of senior girls they were 80 to 93, for 300 employed men, 88 to 95 One study involving retest reliabilities (862) showed even higher reliabilities for 47 graduate students, ranging from 93 to .98 These high reliabilities may be the result of item-transparency and the stability of self-concepts more than of the adequacy of the inventory, Piotrowski's study, mentioned earlier, might be taken as lending support to this interpretation But, whatever it is the Kuder measures, it measures it reliably

Validity Beginning in 1940, and in increasing numbers each year, except for a decline during the last year of the war, studies of the relationship between Kuder scores and other variables have been appearing in the literature According to the writer's count, there was one validation study published in 1940, two in 1941, two in 1942, four in 1943, five in 1944, one in 1945 (by which time publication lag had presumably caught up with the absorption of psychologists in the war effort), three in 1946, and six in 1947 All but one were by persons not directly connected with the inventory, for Kuder has tended to publish his findings only in the manual This demonstrates the recognition on the part of the counselors and psychologists of the need for more evidence concerning the validity of a popular and promising instrument

Intelligence has not frequently been correlated with Kuder scores, perhaps because other problems seemed more vital Adkins and Kuder (8) reported one study of the relationship of interest scores to primary mental abilities, investigation which does have special interest because the mental abilities measured were specific Their data were obtained from 512 university freshmen The correlations between Kuder and PMA Test scores were low, except for one of 39 between number ability and computational interest, a readily understandable relationship Triggs (870) correlated the Kuder with A C E Psychological Examination scores, also finding low correlations, except for one of .40 between literary interests

and verbal scores, and another of .40 between computational interests and quantitative scores, but these relationships held for women only Why they were not found in men, if not merely chance findings in women, is difficult to explain Perhaps social pressure makes college men develop a modicum of computational interest regardless of special ability, whereas women do so only if they have unusual aptitude for such work But this would not explain the relationship between literary interests and verbal ability in women, who are normally both more verbally and more literarily inclined than men More and better studies are needed to clarify these matters

Aptitudes as measured by the Bennett Mechanical Comprehension and Minnesota Paper Form Board Tests were correlated with Kuder scores in a study of 40 aircraft factory foremen by Sartain (671) For the mechanical scale the two correlations were 13 and 15, for the scientific scale 19 and 15, high enough to show some connection, but too low to make the relationship practically important

Interests as measured by Strong's Vocational Interest Blank have been related to Kuder scores in a number of studies, particularly in a series by Triggs (870,871,872,936) Peters (597) first reported correlations ranging from 38 to 52 for 24 college women tested with the Kuder and Strong's Women's Form The correlations between Kuder scientific and Strong physicians' interests, computational and office workers' interests, literary and authors' interests, and social service and lawyers' interests (heavily loaded with the "people" factor in women) were significant, as one would expect So also was that between scientific and lawyers' interests, which is difficult to explain, except on the grounds of their common correlation with intelligence as shown by Strong

Male subjects provided the basis of Triggs' final study (871), in which the trends were similar to those reported for women by Peters For these 166 men the relationships for typical, presumably similar, scales were as given in Table 31

These relationships tend to be what one would expect, but they are low enough so that it would not be possible to use one instrument as a substitute for the other, as many had hoped would be possible On the other hand, the varying degrees of relationship make it possible to use either inventory with better understanding of what is being measured, or both inventories together in order to make a more penetrating analysis of a client's interests

The existence of a higher degree of relationship between the Kuder

scientific and Strong chemist scales (73) than between the mechanical and the chemist scales (51), when contrasted with the inverse order of relationship for the Strong engineer scale (54 and 72), suggests that the Kuder scientific scale assesses a more theoretical, laboratory, or biological type of interest than does the mechanical, and that in testing a would-be engineer it is well to attach more weight to the mechanical scale, while for a would-be chemist the scientific scale should be stressed It is noteworthy that Kuder has revealed an awareness of these relationships in his occupational classification in the manual (446 5–8), for chemist is placed in the scientific group, while the various engineers are placed in the mechanical-scientific It would have been even more accurate, judging

TABLE 31

CORRELATIONS BETWEEN KUDER AND STRONG SCALES

	Kuder						
Strong Scale	*Sci*	*Mech*	*Soc Serv*	*Comput*	*Cler*	*Pers*	*Lit*
Physician	50						
Psychologist	36						
Engineer	54	72					
Chemist	73	51					
Carpenter	26	67					
Math -Sci Teacher	47	46					
YMCA Sec'y			35				
Social Sci Teacher			30				
City School Supt			42				
Accountant				49	55		
Office Worker				25	38		
Life Insurance Sales						58	
Lawyer							50
Author-Journalist							28

by these data, to place the chemists in a scientific-mechanical group (note the order) and leave only the more purely biological occupations in his scientific group

The almost identical correlations between the Strong mathematics-and-science teacher scale on the one hand, and the Kuder scientific and mechanical scales on the other (47 and 46), provide an interesting contrast with both of the sets of relationships discussed in the preceding paragraph, and the closer relationship between the carpenter and mechanical scales as compared with that between the carpenter and scientific scales (67 and .26) further strengthens the interpretation suggested

The clerical scale correlates more closely with both the accounting and the office work scales (55 and 38) than does the computational (49 and 25) This might be taken as a reflection on the computational scale,

but it should be remembered that a good measure of a factor need not necessarily have the most significant relationship to any variable in which it plays a part in other words, the computational factor may be a very real one in some occupations, which actually can be classified in an occupational field in which other factors are more important Accountants do computational work, but they are also concerned with other aspects of office work and record-keeping, as reflected in their clerical interests

The much higher correlation between literary and lawyer (50) than between literary and author-journalist (28) scales is worth noting, for it suggests that the Kuder literary scale is likely to be more valid for legal than for literary occupations Strong's factor analysis of his scales (775 143 and 319) shows that his lawyer and author scales have approximately the same loading of his "things vs people" factor (— 92 and — 98), while the lawyer scale has a slightly heavier loading of the "system" (26 vs — 19) and light loading of the social welfare (— 22 vs — 01) factors It is difficult to rationalize these two sets of data More investigation of the differences between Kuder and Strong scores is clearly needed

Counseling experience has suggested (802) that the apparent discrepancies between Kuder and Strong scores may have diagnostic significance Some persons who made high persuasive scores on Kuder's inventory but low life insurance salesman scores on Strong's seemed on the basis of case history and interview material to be interested in promotional activities, but to dislike activities in which they need to push people to the point of action as in closing a sale The diagnosis and counseling of a number of clients on the basis of this interpretation of differences between persuasive and salesman scores has seemed fruitful, in a few cases even dramatic, but too few have been handled to justify any conclusions It is also possible, for example, that such discrepancies are the result of effects such as that described by Paterson (586), and that the higher Kuder persuasive score is the result of self-delusion or of an attempt to impress the consultant, while the lower Strong salesman score reflects more accurately the true interests of the client If this were the case the selection of salesmen could be improved by using both inventories and devising an index of distortion based on discrepancies between the two scores, the better salesmen would presumably be those whose discrepancy scores were smallest The hypothesis would be worth testing

Personality traits have, we saw in connection with Strong's Blank, generally been assumed to be related to interests This hypothesis was

checked by Evans (243) with the Minnesota Personality Scale and the Minnesota T(hinking), S(ocial), E(motional) Inventory, in relation to the Kuder Preference Record She tested 190 women students at Indiana University, and reported that social introverts tended to score low on the Kuder persuasive interest scale, as did thinking introverts, while extroverts of both types tended to make average or high persuasive scores Thinking extroverts were low also on literary interests, although thinking introverts made average scores on the literary scale Triggs (873) correlated the scores of 35 male and 60 female college students on the Kuder and on the Minnesota Multiphasic Personality Inventory, finding that in men mechanical interests were significantly and *negatively* correlated

TABLE 32

CORRELATIONS BETWEEN SCORES ON THE PREFERENCE RECORD AND ON THE MINNESOTA MULTIPHASIC INVENTORY FOR 35 MALE STUDENTS FROM THE UNIVERSITY OF WASHINGTON FROM TRIGGS (873), UNPUBLISHED PAPER

Scales of the Preference Record	*Scales of the Minnesota Multiphasic Inventory*										
	? 1	*L* 2	*F* 3	*H* 4	*D* 5	H_s 6	P_d 7	M_f 8	P_a 9	P_t 10	S_c 11
1 Mechanical	17	23	− 17	29	03	− 06	− 41*	− 37*	− 18	− 22	− 21
2 Computational	− 11	00	− 22	− 13	− 14	− 15	− 16	− 15	− 42**	− 25	− 18
3 Scientific	− 02	07	− 27	− 14	− 30	15	− 18	− 29	− 36*	− 33*	− 22
4 Persuasive	01	− 22	13	14	21	− 06	08	11	16	23	06
5 Artistic	− 01	− 14	24	09	20	12	21	02	07	19	16
6 Literary	− 15	− 07	14	− 25	07	− 10	19	04	05	07	19
7 Musical	11	− 10	26	− 25	17	22	30	30	25	33*	39*
8 Social Service	01	09	− 27	10	− 34*	15	− 18	17	14	− 16	− 25
9 Clerical	− 01	03	19	− 15	36*	− 24	20	12	01	33*	32*

Level of significance
* 5% = .329
** 1% = .424

with psychopathic and feminine tendencies, computational interests with paranoid, scientific with paranoid and psychasthenic, and social service with depressed tendencies, while musical interests were significantly and *positively* related to psychasthenic and schizophrenic, clerical interests to depressed, psychasthenic, and schizophrenic tendencies Her data are reproduced in Table 32 In women no significant relationships were found between interests and personality traits, although two relationships with validating scores were significant

In view of the currently prevalent idea in guidance centers that social service scores on the Kuder are an indication of personality maladjustment Triggs' findings are especially worthy of note social service interests are shown to accompany wholesome rather than unhealthy personality patterns This does not disprove the observation that some people who want to enter social, educational, or psychological work of

one kind or another are not maladjusted, but it does make one seriously question the tendency to look on high social service scores as indices of maladjustment There is more justification for seeking other signs of disturbance in persons with high musical or clerical scores, but even here the relationships are low enough to make it clear that there are many exceptions Indeed, experience with clients leads the writer to discount musical, artistic, and literary scores unless there is good supporting evidence in the case history, it seems that many people without highly developed interests make high scores on one or more of these scales, presumably because most high school and college graduates enjoy listening to some kind of music, looking at some kinds of pictures, and reading fiction enough to seem interested in one of these fields if other more definite interests are lacking

Grades or other indices of academic achievement have been correlated with Kuder scores in at least ten studies Triggs (870) found correlations of 42 (women) and 32 (men) between scientific interests and general science achievement, 40 (men) and 33 (women) between literary interests and achievement in English literature, 34 (men) and 36 (women) between computational interests and mathematical scores Yum (952) found significant relationships between the literary interests and grades of men (335) and between the computational interests and average grades of women (295) at the University of Chicago, but the comparable relationships for the opposite sex were in each case not significant Crosby (184) reported significant differences between the chemistry and biology grades of high- and low-scoring scientific interest groups (critical ratios = 7 6 and 12 2), and between the accounting grades of high- and low-scoring computational interest groups (6 9) The 1946 manual cites a thesis by Mangold (506), in which she found significant relationships between scientific interests and scores on the co-operative Natural Science Test (385), literary interests and Co-operative English Test scores (31) and literary interests and literary scores on the Co-operative Contemporary Affairs Test (59) Detchen (199) developed a scale based on 109 of the 785 Kuder items which were found to differentiate A and B students from D and E students, and obtained a validity coefficient of 60 with a social science comprehensive examination as her criterion, her subjects were 247 students in the original group, 106 in the cross-validation group for whom the validity coefficient shrank to a still significant 55 The typewriting and stenography grades of women liberal arts students, 96 and 75 in number, were related to Kuder clerical scores by Barrett (45),

who found that the interest scores did differentiate superior (A and B) stenography students from inferior (D and F) students, the cut-off score being the 55th percentile, the scale had no validity for typing Dentistry grades were the criterion used by Thompson (825), he found no relationship (— 06) between mechanical interests and dental practicum, while the validity of the social service scale was 24 These seemingly odd results may perhaps be explained by the very high mean mechanical interest scores (91st percentile) and their restricted range, whereas the social service scores had a lower mean (67th percentile) and presumably a greater range On the other hand, scientific interests correlated 28 with theory grades, as anticipated

Achievement on the USAFI Tests of General Educational Development was related to Kuder scores in a well designed study by Frandsen (271) Achievement in the natural sciences correlated 31 with computational and 50 with scientific interests, in the social studies, — 37 with social service but 31 with literary and 34 with scientific interests, probably because of the respectively negative and positive correlations between those types of interests and academic ability Frandsen cites a master's thesis in which Turner reported correlations of 29 and 32 between scientific interests and grade point-ratio in several courses in the biological and physical sciences, and 49 between computational interest and grades in physical sciences On the basis of his and other findings Frandsen appropriately concluded that "science and mathematical interests are definitely related to general achievement in parallel areas For other areas, significant and logically consistent interest-achievement relationships have not been so clearly indicated, though some slight relationships have been noted for literature and social studies' Exceptions, Frandsen goes on to state, appear to be due to more fundamental negative relationships between social service interests and mental ability

Completion of Training From this point Frandsen proceeded to check Strong's hypothesis that interest would result in remaining in rather than leaving a field of endeavor, by correlating Kuder scores with percent of total credit in scientific and social studies The correlations are shown in Table 33

These data support Strong s hypothesis students with social service interests tend to choose more social studies courses, and students with scientific interests tend to elect more scientific courses Further confirmation is found in a study by Bolanovich and Goodman (109), in which the engineering grades of 66 women students of electronics in the Radio

Corporation of America's war-training program correlated only 09, 18, and 10 with mechanical, computational, and scientific interests on the Kuder, but the scientific and computational interests of the cadettes who successfully completed training were significantly higher than those of women who did not complete it, while those who were released scored significantly higher than others on the persuasive scale. These two studies seem to provide convincing evidence that what Strong found with his inventory is also true of the educational predictive value of Kuder's

Occupational choice was related to Kuder scores by Crosby and Winsor (185), by Kopp and Tussing (441), and Rose (647) The first authors asked college students to estimate their interest in the seven types of activities measured by the then current form of the Preference Record, and correlated these estimates with scores on the interest inventory, the average coefficient was 54, and there was more agreement between the two indices for the more intelligent (as measured by the A C E) than for

TABLE 33

CORRELATIONS BETWEEN KUDER SCORES AND CHOICE OF COURSES

	Percent Total Credit in	
Kuder Interest Scale	*Natural Sciences*	*Social Studies*
Scientific	54	− 35
Social Service	− 17	32

the less intelligent students Kopp and Tussing found similar results with approximately 50 high school boys and an equal number of high school girls (r = 59 and 50), using the nine categories of the revised Preference Record Rose used a similar procedure with 60 veterans, finding a correlation of 61 between inventoried and expressed preferences Those who had specific objectives showed no closer agreement than others About two-thirds of the group preferred occupations in fields in which they made high scores These results are consistent with those already seen for Strong's Blank

Success in an occupation has been correlated with scores on the Kuder in only three published studies at the time of writing In the first of these, Sartain (671) administered a battery of tests to 40 foremen and assistant foremen in an aircraft factory who were rated by their supervisors The ratings had an interform reliability of 79, but yielded significant correlations with none of the instruments, that for the Kuder mechanical scale was 07, social service scale − 06, and clerical scale 003 In the second study, initiated by the writer and reported by Guilford (316 613–616), the Kuder was administered to 937 AAF pilot cadets who later took

primary training The correlations with success in training were statistically significant for only one scale, and that coefficient was only −.10, between social science interests and success The validity coefficient for the mechanical scale was only 02, and the musical and artistic scales, which on the basis of results from information tests and biographical data blanks would be expected to have negative validities, actually had low but nearly significant positive validities (05 and 08) Guilford suggests that this is because the Kuder scales sample interest and appreciation, whereas the more valid (for predicting success) tests of information and biographical data sample experience Thompson (826) found superior management engineer executives more interested in mechanical and less interested in social service activities than average men in similar jobs

In an unpublished study, reported in an abstract of a paper by Di-Michael and Dabelstein (200), efficiency ratings of 100 vocational rehabilitation workers were correlated with Kuder scales Of 48 relationships computed, the first two of those which follow were significant at the one percent level, the third at the 5 percent level

r promotional work and persuasive interest score = 32,
r professional reading and scientific interest score = 26,
r employer contacts and persuasive interest score = 19

These findings suggest that although uncorrelated with overall success in a job, interest as measured by the Kuder may be related to success in some aspects or duties of a varied job

Occupational differentiation on the basis of Kuder scores has been most extensively reported in the manual, in which Kuder reports patterns for a number of men's and women s occupational groups These have already been discussed in connection with the norming of the Preference Record, it was pointed out that the numbers in each field are distressingly small, and the fact that the selection of the samples of each occupation is not made clear suggests that it was opportunistic rather than planned It has also been seen that in the case of one women's occupation, nursing, increasing the size of the sample made little difference in the mean or standard deviation Brief verbal summaries of the patterns revealed in Kuder's work are given below, as a tentative guide to the interpretation of the scores

Men in social welfare occupations, e g , vocational rehabilitation supervisors, clergymen, social workers, school administrators, and teachers of social studies in high schools, tend to make high social service and

literary scores, personnel managers, however, are somewhat less distinguished by high scores on these areas, and unlike the social welfare group tend to make equally high scores on the persuasive scale

Men in literary occupations, such as writers, English teachers, and actors, tend to make high literary and musical scores, but actors are also high in artistic interests, lawyers and judges differ even more in that they make high scores in the persuasive area as well as in the literary and musical

Scientists such as chemists and engineers tend to make high scores on the scientific scale, electrical and especially industrial and mechanical engineers also making high scores on the mechanical scale The computational scores of these groups are higher than average, but only in the case of the industrial engineers are they significantly high The only significantly high score made by the 26 draftsmen was in the artistic area Spear (730) found similar trends in engineering freshmen, as did Baggaley (36) with liberal arts college freshmen

Clerical workers, including accountants, auditors, bookkeepers, and cashiers tend to make high computational and clerical scores, the higher-level groups being outstanding in computational and the lower-level groups in clerical interests

Salesmen and sales managers make their highest scores on the persuasive scale, this being the only outstanding score of salesmen who sell to individual consumers, while those who sell to distributors or manufacturers tend also to make high clerical interest scores Judging by pattern inspection, life insurance agents ($N = 24$) do not differ appreciably from other salesmen, a finding which is at variance with Strong's data, previously discussed

The patterns for women are in most cases similar to those of men in the same field, and like the men's, they tend to agree with expectation Women physicians tend to make high scores in scientific and mechanical fields, as do laboratory technicians, but neither group is high in computational interests (no similar men's groups were tested) Nurses make their high scores in scientific and social service areas, but it is noteworthy, in view of Strong's findings concerning some women's occupations, that none of the means are as high as the 75th percentile in other words, they are a relatively undifferentiated group This is true also of women telephone operators, stenographers and typists, teachers of home economics, and teachers of social studies, as Strong's work would lead one to expect

Groups of 50 male life insurance salesmen and 50 social workers were

tested by Lewis (469), who found the former significantly higher than the general population in persuasive and the latter significantly higher in social service interests Profile analysis was not made, however Lehman (465) followed up students of home economics at Ohio State University, finding from 10 to 125 in each of several subdivisions of the field Teachers, the largest group, scored high on social service, artistic, and scientific interests, hospital dieticians were high in social service, scientific and computational areas, restaurant and tea room managers scored high on the artistic and computational scales, home service and equipment workers made high scores in social service and persuasive fields, and journalists in the literary and artistic fields Women marines were tested by Hahn and Williams (323), who found relationships between interest patterns and duty assignments which, like those just reviewed, were in line with expectation

Job satisfaction has so far been used as a criterion only by Hahn and Williams, in the study just referred to and by DiMichael and Dabelstein (200) The former found that satisfied clerical workers were significantly more interested in clerical activities as measured by the Kuder than were dissatisfied clerical workers, the critical ratios for three sub-groups being 2 28, 2 41, and 2 97 Clerk-typists who were dissatisfied tended to be more interested in mechanical matters, general clerks who were satisfied were also more interested in computational activities

DiMichael and Dabelstein (200) correlated satisfaction with various job duties, as rated by 100 vocational rehabilitation counselors, with scores on appropriate Kuder scales administered five months previously The correlation between enjoying "contacting employers to secure jobs" with the Kuder persuasive scale was 28, and between "handling clerical details" and clerical interest scores 32 None of the expected relationships between social service aspects of the job and social service interests were significantly correlated for this group Another group of 46 male counselors were tested *after* they had made the job satisfaction ratings, and it is interesting that here the correlation between enjoyment of the job as a whole and social service interest score was 29 as opposed to .13 for the other group, that between enjoying interviewing clients and social service interest score rose from 06 to 43, and other expected relationships became closer This might be attributed to either of two factors The first group may have lacked insight into their interests when they filled out the Preference Record, the subsequently completed satisfaction questionnaire therefore being a more accurate picture of their interests

Or the second group, having filled out the satisfaction questionnaire first, may have answered the interest inventory more searchingly and insightfully, perhaps even distorting answers in order to make them consistent with what they had already said As the first group had had 2 5 years experience on the job, and the second only 1 year, it does not seem likely that the first explanation is correct The first group knew its work, but did not know it was going to rate it for job satisfaction, the second group also knew its work, though less well, and had already rated it for satisfaction The closer agreement between the two indices in the latter group must therefore be related to having job satisfaction in mind when they took the Preference Record It would be interesting to know to what extent the greater agreement represents, respectively, stereotyping, insight, and distortion

Use of the Kuder Preference Record in Counseling and Selection It has been established that the traits measured by the Kuder are internally consistent and relatively independent of each other They are not closely related to intelligence, although there appears to be a degree of relationship between some primary mental abilities and the expected interests Similarly, special aptitudes such as mechanical comprehension seem to be somewhat related to appropriate interests The relationships between Kuder and Strong scores are found according to expectation, but they are not high enough to justify using Kuder scores as though they were obtained from Strong's Blank The reason for this is obvious enough the Kuder scores measure relatively pure interest factors, whereas Strong scores measure the interests of people in occupations Chemists, for example, are characterized by interests which are partly scientific and partly mechanical, while mechanical engineers have a combination of mechanical, scientific, and computational interests Personality traits have also been found to be related, in some instances, to interests as measured by the Preference Record contrary to a commonly held opinion among vocational counselors and psychologists in guidance centers, interest in social service is related to wholesome personality patterns as measured by the Minnesota Multiphasic Personality Inventory, as are mechanical interests, on the other hand, the personality patterns associated with musical and clerical interests are not so healthy

The development of interests as measured by Kuder's inventory is not clear Data so far collected indicate that there are no significant changes associated with age during high school and college years, but as this tentative finding is contradicted by the much more intensive and exten-

sive work with Strong's inventory it seems well not to draw any conclusions until after more thorough-going studies are completed

Occupational significance of scores on the Preference Record has been demonstrated largely by compilation of means and sigmas for people employed in various occupations Although the numbers are small, the data indicates differences between groups such as would be hypothesized on the basis of Strong's results The development of occupational indices, or procedures for the statistical comparison of an individual's scores with those of people in various occupations will make the occupational interpretation of the Kuder more objective, but it will take some time to make an appreciable number of these available In the meantime we have seen reason for thinking that Kuder's classification of occupations by interest types has some validity, although the published materials indicate that as yet much of the classification has no empirical basis The little material available on the relationship between Kuder scores and success on the job is less encouraging than for Strong's inventory, although one study has shown some relationship between interest and success in appropriate duties or aspects of the job

In schools and colleges the Kuder does seem to have real possibilities even for the prediction of success in courses, for scores are significantly related not only to the completion of training, as for Strong's Blank, but also to grades in some appropriate subjects, specifically the scientific and mathematical Validity for other subjects is more doubtful, at least when the interest-range is as restricted as it generally is The scorability of this inventory, the ease with which student participation in scoring, converting scores, and plotting profiles lends itself to interpretation of results and discussion of their implications, give the Kuder many advantages for use in school and college guidance programs Its transparency is presumably less important in counseling than in selection programs, and the fact that scores have only moderately high correlations with expressed preferences shows that it can contribute something to the diagnosis of interests, especially for the least able students for whom the discrepancy between choices and scores is greatest

In guidance centers, whose clients are generally somewhat more mature and more experienced than students, it is especially desirable to make a careful study of the manifest interests of clients to whom the Kuder is administered, as a precaution against overemphasis on the literary, musical, and artistic scores which seem often to be high simply on an appreciation basis Even in schools this can be similarly checked, but there the

counselor may need, and be able, to depend partly on try-out experiences Differences between Kuder and Strong scores often suggest new interpretations worth exploring in interviews, making the use of both instruments desirable in difficult cases

The value of the Kuder in employment centers and in business and industry is still virtually unknown, as it has been little used in such situations Despite its "industrial" short form it was apparently not designed with such use in mind, and its transparency has militated against it For it to be valuable in personnel selection or evaluation programs more research should be done, including studies of the extent of faking among applicants, the possibility of a distortion score and the development of occupational indices appropriate to the jobs of the specific company or institution

The Allport-Vernon Study of Values (Houghton-Mifflin, 1931)

This inventory was developed by G W Allport and P E Vernon in an attempt to measure the personality traits postulated by Spranger in his *Types of Men* (734) The traits measured are best described as values or evaluative attitudes, although some of them verge on needs (see next chapter) We have seen that they closely resemble interests but are perhaps correctly described as more basic, for they concern the valuation of all types of activities and goals, and they seem in some instances to be more closely related to needs or drives In practice, however, values and interest inventories are often used more or less interchangeably, and their relationships warrant treating them as interest inventories The Allport-Vernon is by no means the only values test, but it is the first of its kind, has been the most thoroughly studied, and is still the most widely used A review of work with this and other values tests was published in 1940 by Duffy (216)

Applicability The Study of Values was designed for use with college students, and more as an instrument for research in the theory and organization of personality than as a practical aid in counseling or selection Its vocabulary level is therefore higher than that of most inventories, Stefflre (752) has shown that it has a vocabulary grade placement of 11 3, and that only the Cleeton Vocational Interest Inventory, among the widely used blanks, is more difficult to comprehend For these reasons the Allport-Vernon should be used only with superior high school juniors or seniors, college students, or superior adults Even for these some of the items may be difficult to accept, if not to understand, because of their

seemingly esoteric nature College students usually take them in their stride, but employment applicants are often impatient with some of the mystical and aesthetic items

Changes in scores during the four years in college have been studied by Harris (342), Schaefer (674), and Whitely (922), and summarized by Duffy (216) as showing that " the lowest coefficients of correlation are found always between the first and other administrations of the test, and that the trend (perhaps not statistically significant) is toward an increase in aesthetic, social, and theoretical values, and a decrease in religious, political, and economic values " Subsequent studies by Arsenian (32) and Burgemeister (124) with college men and women do not alter these conclusions, which fit in with conclusions concerning the increase in social welfare interests with age in adolescence, but contradict other data on scientific interests, and have no counterpart in so far as aesthetic and other values are concerned It may be that the increases in aesthetic and theoretical interests, and decreases in religious and other values, are the result not of maturation but rather of college experiences It would be helpful to have retest data for these same persons five and fifteen years after graduation from college, but none are available Neither are there studies of age changes in other more typical populations

Content The Allport-Vernon consists of 45 items, the first 30 of which are paired comparisons and the last 15 multiple-choice, making 120 alternatives in all As in the Kuder Preference Record each of the choices represents one of the types of interests or values and the corrected sum of the examinee's choices of any one kind of item constitutes his score for that type of value As in the Kuder, a higher score on one type of value automatically makes for a lower score on some other type or types The items are designed to tap theoretical (interest in truth and knowledge), economic (interest in the useful or material) aesthetic (interest in form and harmony), social (interest in social welfare), political (interest in prestige and power), and religious (described as interest in unity with the cosmos but actually adherence to the forms of religion) values The use of Spranger's esoteric terminology has created many misunderstandings of the traits measured, not only among users of the test but also in some investigators who have taken the terms in their common rather than very special sense Even Spranger's definitions are misleading, as just noted in the case of religious values, because of poor implementation of the authors' intentions The writer has frequently noted, for example, that high school students from traditionally religious homes, in

whom observation and study revealed no real depth of religious feeling or belief, make high religious scores on the Study of Values In their cases the scale seems to measure only verbal conformity to formal religion Any user of the inventory should therefore study the items carefully, as well as the authors definitions, before making interpretations

Administration and Scoring The blank requires from 20 to 40 minutes to administer, depending upon the verbal ability of the examinee There is no actual time limit, but rapid work should be encouraged Directions are simple and clear Scoring is by means of a self-explaining scoring and profile sheet, readily understood by college students Final raw scores may be converted into deciles by a table on the profile sheet, but the small number of items makes the conversion very crude and complicates interpretation The plotting of final raw scores on the profile sheet brings out the dominant values more effectively and with less exaggeration, and is to be recommended for use This scoring procedure is more time-consuming than most in current use, but this is of minor importance when the inventory is used as a part of class work and is scored by the students The use of the profile is helpful in stimulating discussions of values and goals, and in bringing about self-insight

Norms The college student norms provided by the manual have been found reasonably adequate in a number of studies (342,136), with variations which seem explainable in terms of the clientele and emphasis of the colleges in question But these norms are general, and serve, like Kuder's, only as a backdrop against which to study the variations of part scores within an individual Occupational norms are also desirable, in order to throw light on the vocational significance of the scale, but are not available except for 26 YWCA secretaries (17) On the other hand, the mean scores made by a great variety of college curricular or pre-occupational groups have been reported in various studies referred to below in the section on occupational differences These lend support to the practice of interpreting Allport-Vernon scores in vocational terms

Standardization and Initial Validation The diagnostic efficiency of the inventory was tested by the internal consistency method in the original study (898), in which it was found that the scales were relatively reliable and independent, only the social values scale being of questionable reliability (65) Scores correlated 53 with students' self-ratings on similar traits on the average (range of r's = − 06 to 69), even though the reliability of the ratings was only 59, suggesting consistency between most self-concepts and self-described behavior The one low intercorrela-

tion was for social values Expected differences were found between curricular groups, science majors, for example, being high on theoretical and low on economic values, while business students tended to score high on economic values

Reliability As previously noted, the reliability of the social values scale was found to be only 65 (898), but the average retest reliability after three weeks was 82, showing considerable stability in the other scores, these findings of the original authors have since been confirmed by other investigators (136)

Validity Scores on the Allport-Vernon have been related to most of the variables which can be studied in college populations, although to relatively few which are observable only in other groups

Intelligence test scores have been correlated with values scores, for example, by Pintner (606) in a study of 53 graduate students of educational psychology, for whom the correlations were 24 with theoretical, 38 with social, — 28 with political, and — 41 with economic values, those with other values being practically zero Other studies, summarized in the manual, in Cantril and Allport (136), and in Duffy (216) reveal similar trends except for social values, the results for which are generally not so clearly positive

Grades were used as a criterion in Pintner's study (606), but as they were based partly on performance in test administration they are somewhat atypical social values correlated 46 with grades, while the other coefficients were so small as to be negligible Cantril and Allport (136) found theoretical values correlated with sociology grades at Dartmouth to the extent of 25 In a study of students at Sarah Lawrence College, Duffy and Crissy (217) found a validity of 34 for a combination of values scores, using ratings of academic achievement at the end of the freshman year as their criterion Theoretical and aesthetic values had positive weights, economic and political negative With the Co-operative Test of General Culture as a criterion, Schaefer (674) found relationships of 58 and — 47 between the literary achievement and aesthetic and economic values of 51 women sophomores, 47 and — 28 between fine arts and aesthetic and economic values, 37 and — 37 between history and the same values, and 31 between general science and theoretical value These relationships seem unduly high, and may be peculiar to the local situation (Reed College), they would in any case need confirmation before being applied elsewhere A safe generalization from the studies reviewed would seem to be that there is a slight tendency for students

with theoretical values to make better grades than students in whom other values are dominant, a conclusion which is congruent with the definition of the trait, and that in some situations other values will be associated with success in appropriate fields of endeavor

Success on the job has not, to this writer's knowledge, been related to scores on the Allport-Vernon Study of Values

Occupational differences have not been studied by means of employed men or women, but numerous studies have shown that professional students are differentiated by the Study of Values in accordance with expectations *Theoretical* values are found in students of education (342), engineering (342), medicine (342,763), natural science (674), and social studies (674) *Economic* values characterize only students of business (763,874) *Aesthetic* values are strong in students of drama (293), education (763), literature (674,763), and the social studies (674) *Social* values have not so frequently been studied, as the scale is not reliable enough for individual diagnosis, it is adequate for the study of group trends, which show that YWCA secretaries (17) stand high on it, but, surprisingly, that students majoring in the social studies (674) tend to make low scores *Political* values are significantly high in engineering students (342), physical education students (695), and law students (342,763) *Religious* values have been found to be high in seminarians (492) and in YWCA secretaries (17), but the high scores of high school commercial students (874) and low scores of college students of business (763) suggest that the religious values scores do not, in some cases, represent more than the lip service of immature persons who have as yet experienced neither deep religious feeling nor intellectual doubts concerning religion

Satisfaction in one's work has not been related to scores on the Study of Values, as might be expected in view of its limited occupational use

Use of the Allport-Vernon Study of Values in Counseling and Selection The traits measured by this inventory resemble those measured by the other inventories studied in this chapter Like the Kuder, it taps interest factors which play a part in a variety of occupational fields, usually in ways which would be anticipated in view of the nature of the items However, the traits appear to be somewhat more fundamental and more closely related to basic needs and drives than those measured by other interest inventories They have been found to change somewhat during the college years, social interests increasing as other studies have also reported, but increases in theoretical and aesthetic values may be related to specific college influences, together with decreases in religious

and economic values Too little is known concerning age changes in values Values are related to intelligence in the same way as interests

Occupations for which the Study of Values has significance appear to be largely at the professional and executive levels, but that is due to the vocabulary and intended use of the instrument Values are related in expected ways to choice of training in fields such as art, business, drama, education, engineering, law, literature, medicine, natural science, psychology, the priesthood, social studies, and social work Only in the last-named field have experienced workers been tested, but the data for training groups are consistent enough to justify some confidence in their occupational significance As no norms are available, the counselor must interpret on the basis of peaks and valleys in the profile, a procedure which is safer with this instrument than with most when drawing conclusions from high scores because of the method of construction, but more dangerous with low scores or valleys in the profile because interest in such a field may be very strong even though pressed down artificially in the mutually-exclusive response technique

In schools and colleges this inventory may have some value in determining appropriate fields in which to major although it generally has less value for predicting grades than an intelligence test The nature and degree of the relationships between values and grades in various types of courses are likely to vary with the institution, because of the importance of climates of opinion in attracting students and in modifying values Differences in predominant values or climates of opinion in different colleges give the test some value in helping students choose congenial colleges The self-scoring feature of the inventory makes its use in orientation and psychology classes easy, and it lends itself well to the starting of discussions of values, interests, and vocational objectives, such as is appropriate to orientation programs The esoteric nature of some of the items limits its usefulness, however, to moderately well motivated persons, and the vocabulary limits it to superior high school and to college students

In guidance centers the Study of Values can be helpful in aiding potential college students in the choice of colleges in which they will find the psychological atmosphere congenial and conducive to growth, although for this purpose comparisons between the mean scores of students in different colleges need to be made more systematically than has so far been done A survey of the literature with this purpose in mind would yield some useful material More important than this use, in

guidance centers, is the diagnosis of interests when it is suspected that Kuder or Strong scores are distorted by a clear-cut but inappropriate self-concept The non-vocational nature of the Allport-Vernon items presumably makes them less subject to choice on the basis of vocational stereotypes, and more on their own merits, than the more clearly occupational items in the Kuder and even the non-occupational parts of the Strong Unfortunately this hypothesis has never been tested Until it is, the clinical counselor in search of an understanding of a puzzling client cannot afford to neglect this test and so to miss the chance to sink a shaft into the interest field which is slightly different from those sunk by other instruments

In employment services, business, and industry this inventory is likely to be less useful than in other types of counseling or selection situations The vocabulary and subject-matter make it seem out-of-place to employment applicants, and the norms and validation do not lend themselves to as effective use in selection programs as do those of certain other interest inventories An industrial and business version might presumably be constructed and be of considerable value in selection because of the differences between it and the standard vocational interest inventories But such a project has yet to be planned and carried out

The Cleeton Vocational Interest Inventory (McKnight and McKnight, 1937, 1943)

This inventory appears to have been developed in an attempt to simplify the scoring of Strong's Vocational Interest Blank, and incorporates many items used in it and in other inventories constructed in the Carnegie tradition It has been rather widely used in schools, colleges, and guidance centers, but has not enjoyed the popularity of either the Strong, despite its simpler scoring, or the Kuder, which captured a large segment of the vocational-test-using public almost on publication The writer believes that this may be due partly to warranted misgivings concerning the transparency of items grouped according to their occupational significance, and partly to such an irrational thing as dislike of the meaningless and difficult-to-remember codes used to designate the occupational families Whether scientific or not, convenient handles help

Description The Cleeton Inventory was designed for use in grades 9 through college, and with adults, but was constructed on the latter and has a vocabulary grade placement of 12 (752), making it the most difficult of the well-known interest inventories Both men's and women's forms

consist of ten groups of items, each group representing an occupational family (e g , OCA clerks, stenographers, typists, and other office work occupations) and consisting of 70 items, 30 of which are occupational titles, 20 names of school subjects, magazines, prominent persons, etc , and 20 leisure-time activities, work activities, and peculiarities of people. Scoring is done by adding unitary weights for each item marked in a given group It was standardized by administering it to some 7000 individuals engaged in a variety of occupations, principally in the Pittsburgh area In 76 percent of 1,741 cases the highest inventory rating agreed with the occupation engaged in, while in 95 percent one of the three highest ranking groups included the occupation engaged in

The scores are quite reliable, ranging from about 82 to about 91 (manual) However, as many have pointed out (622), the grouping of items by occupational families makes them easily recognizable and spuriously increases reliability an examinee readily sees that a given section is, e g , the engineering section, reacts "I want to be an engineer, I like these," and gives favorable responses to some items which would be marked differently if they were scattered among other items Unfortunately this hypothesis has not been checked experimentally, but counseling practice suggests, and the relatively high reliabilities seem to confirm the hypothesis, that this is a valid criticism

Validity There have been few studies of the validity of the Cleeton, further testimony of the fact that it has not challenged most vocational psychologists, most of the published studies are not concerned with the relationship between inventory scores and external criteria It was administered to students of education by Congdon (168), who found significant differences between men and women who planned to teach, on the one hand, and who planned not to teach, on the other She also found that scores in the field of claimed interest were higher than scores in fields in which no interest was claimed, but this is not surprising in an inventory as seemingly transparent as this Even the former finding may be spuriously high because of the same sort of halo effect or stereotyping

The correlations between Cleeton scores and Strong's scales were computed by Arsenian (29) for 150 Springfield College freshmen who took the two inventories at intervals of one week (Strong Blank first) Scores for the Strong scales which belong to the same occupational family were combined to yield group scores comparable to Cleeton's, and the two sets were correlated The coefficients of correlation ranged from 16 (LFJ and Lawyer-Author-Journalist) to 68 (TMD and the social

welfare scales), the average being .45 This is slightly lower than the correlations between the Strong and the Kuder, which have less item-similarity than the Strong and the Cleeton It would clearly not be wise to use the Cleeton as a substitute for the Strong Blank, although there is considerable similarity in the inventories and in the meaning of the scores

Use of the Cleeton Vocational Interest Inventory. In view of the availability of more thoroughly studied inventories such as the Strong, the Allport-Vernon, and, more recently, the Kuder, there is little justification for using an instrument concerning which there is still so much room for questioning and for which there is still little in the way of field validation Although Cleeton's standardization data are rather impressive, there has not yet been enough follow-through on the inventory to make it a well-understood instrument

The Lee-Thorpe Occupational Interest Inventory (California Test Bureau, 1943)

This new inventory has been available for so short a time that practically nothing has appeared concerning it in the professional journals The writer has located no studies of its validity, and practically all that is known concerning it is in the manual and "Occupational Selection Aid" supplied with it

Description The items were written in simple language, with a vocabulary grade placement of only 6 8 (752), it (Advanced Form A) is therefore easily understood by junior and senior high school boys and girls The paired comparison form is easily handled also at that level The items are not, however, offensive to adults, they are based on the *Dictionary of Occupational Titles* (888), and so have the aura of authenticity It is scored for fields somewhat like Kuder's, by simple item-count The inventory itself therefore looks attractive to users of vocational tests The manual shows that it is reliable (71 to 93) The norms are based on 1000 12th-grade students, and are said to be applicable to any high school grade and to adults—a fact which seems improbable, in view of Strong and Carter's work and of tentative findings reported by Lindgren (473)

Validity The only claims for validity set forth by the manual are based on the source of items, the design of the items, the balance of activities sampled, and the presentation of items All of these, it should be noted, are internal, not external, criteria, and are dependent upon

the good judgment of the test authors rather than upon objective evidence. The inventory is therefore still in the embryonic stages, lacking evidence of occupational validity Lindgren (473) has, however, reported a substantial relationship between appropriate Lee-Thorpe and Kuder scales

Use of the Lee-Thorpe Occupational Interest Inventory The nature of the inventory makes it attractive to potential users, but it is at present a purely experimental form which has yet to be validated against occupational criteria. It may therefore be used in research by those who have the resources for conducting validation studies, or as an interview aid, but has no value at this point as a diagnostic or prognostic instrument

The Michigan Vocabulary Profile Test (World Book Co, 1939)

Unlike the other instruments discussed in this chapter, this is a test rather than an inventory It is virtually the one information test of interests now available, although the Army Air Forces (316 Ch 14) developed one which was quite valid for pilot and navigator selection and will no doubt stimulate civilian counterparts The Michigan test was developed by E B Greene, as a test of specialized vocabulary which might be prognostic of interest and success in several fields of activity It was little used before World War II, but has since been widely used in work with veterans

Description Two forms are available, each of which was designed for high school and college use and has eight divisions human relations, commerce, government, physical sciences, biological sciences, mathematics, fine arts, and sports There are 240 items divided among these eight areas, each phrased as a definition followed by four terms from which the one which corresponds to the definition must be selected Items are arranged in ten levels of difficulty, three items per level An attempt was made to eliminate terms which could be guessed by knowledge of roots, prefixes, etc, thus reducing the effects of reasoning and restricting the test to information The items were selected from more than 6000 submitted by students in the various fields Groups of items were refined by internal consistency analysis, all items being required to correlate 30 or above with the score on that part The inter-form reliabilities range from 78 to 94, with a median of 81 Administration is untimed, most college students finishing in about one hour and high school students sometimes requiring as much as one and one-half hours The test can be machine or hand-scored with stencils, the score is the

number right for each part A profile chart is provided on the answer sheet Norms are expressed in percentiles, and are based on 4677 students from 9th grade through college, and are available for both part and total scores, this means that each norm group contains an average of slightly less than 600 persons Because of the limited number of items in each scale, the percentiles change rapidly a raw score of 16 on the human relations scale places a college freshman at the 31st percentile, while one of 17 places him at the 50th This is the unfortunate result of a steeply graded test, it would probably have been better to have separate forms for high school and college, and to have more items working at each level in order to get a better spread of raw scores and of percentiles As it is, too much emphasis is put upon chance factors which affect the answering of any one item Increases in scores with grade occur, as would be expected in a vocabulary test Finally, profiles are given for students in several professional curricula, including law, nursing, engineering, business administration, medicine, education, and social studies, the numbers for these groups ranging from 125 to 182 These do not actually constitute norms, as only the means are given, but they do aid in interpretation

Validity Unfortunately there have been almost no studies of the relationship between scores on the Michigan Vocabulary Profile Test and other variables, although data are needed on the relationships with intelligence, inventoried interests, grades, completion of training, occupational choice, success in various occupational fields, job satisfaction, and other external criteria It is surprising that an instrument which has been as widely used as this during the postwar years has had so little publication, presumably this deficiency will be remedied after sufficient time has elapsed for analysis of the data accumulated by the veterans' testing and counseling programs One bit of internal evidence concerning the validity of the test is contained in the manual, which shows that none of the part scores correlate more than .54 with any other, the averages for each scale ranging from .15 to .34 Thompson (826) has reported differences between more and less successful executives

Use of the Michigan Vocabulary Profile Test Like many other published tests this one is still in an embryonic stage because there has been no follow through in the collection and publication of validation data and vocational norms It has been widely used since World War II in work with veterans, because its grade norms for specialized vocabularies have made easier the evaluation of the readiness to resume a high

school or college education of somewhat mature young men whose education had been interrupted Clients whose informal education has given them much of the vocabulary of a special field can be assumed (there is no actual published evidence) to have some of the prerequisites of success in that field The usefulness of the Michigan Vocabulary Profile Test will probably be limited to such cases, and to the diagnosis of reasons for failure in educational programs, until more complete validation has been carried through

Trends in New Measures of Interests

Although the discussion of the widely used measures of interest which constitutes the body of this chapter has brought out many of the important trends in interest test construction, there are certain other developments which are not made clear by work with these instruments For this reason developments with some less widely used tests, some of them not available for general use, are briefly considered in closing this chapter

The use of simpler and more familiar items, describing or pertaining to activities which have been almost certainly within sight and reach of the subjects for whom the instrument is designed, is one trend which seems clear in recent interest inventories We have seen that the Lee-Thorpe inventory succeeded in keeping a 6th grade vocabulary level The *Dunlap Academic Preference Blank* (World Book Co , 1939) was developed for use in grades six through nine, utilizing vocabulary items related to the subject matter of those grades and familiar to pupils through their studies (219,220,713), it yields scores for degree of interest in literature, geography, arithmetic, history and other subject areas, plus measures of mental ability The *Gregory Academic Interest Inventory* (Sheridan Supply Co , 1947) is a somewhat similar inventory, based on liking for high school subjects and activities (312), and designed to help college students in the selection of challenging curricula An *Activities Interest Inventory* on which T L. Kelley has worked for some years (562) attempts to tap only activities with which the typical respondent (high school youth and wartime Army enlisted men in some of the basic studies) is familiar without occupational experience and to use only terms easily understood by him In so far as it insures comprehension by the subject and uniformity of interpretation this is a highly desirable trend, but if, as it seems may have been the case with the Kuder, this increases the transparency of the inventory to the point of risking its

essential validity as a measure of underlying interests, this would be unfortunate This is not a necessary resultant, however, as it should be possible to locate an ample number of items the meaning of which is clear to the subjects for whom the inventory is intended, but the occupational significance of which is hidden Strong's Vocational Interest Blank seems to contain a number of these

The measurement of factors appears to be favored in the development of new inventories, rather than the measurement of interests peculiar to specified occupations To some extent this is a reflection of current interest in factor analysis, and perhaps even of a realization of the contribution which factor analysis can make to the purification of measures and the improvement of predictions, as pointed out by Guilford (316, 317) But as most of the inventories which measure types of interests (interest "factors") have arrived at these by methods other than factor analysis, however legitimate, and have not developed occupational norms to serve as a guide in the interpretation of the factor scores (Kuder shows signs of becoming a notable exception to this generalization), one is inclined to suspect that the trend is in part the result of a tendency to choose the easy and the short way, to rely on *a priori* or at best internal indices of occupational significance rather than on external criteria Test constructors and users should therefore be wary of the interest inventory which measures types of interests without providing objective evidence of the occupational significance of these interest factors

Information tests of interests are again gaining favor, as factor analysis and related internal-consistency and item-validation techniques are making it possible to construct instruments which measure information important to a variety of fields in a reasonable length of time For example, it takes the O'Rourke Mechanical Aptitude Test, one of the first information tests and interest and aptitude, nearly one hour to measure mechanical information, whereas the Air Forces' General Information Test assessed interests of differential significance for success as bombardier, navigator, and pilot in no greater length of time A few words on the nature of the instruments may be worthwhile, in order to make clearer the direction developments may take

The AAF General Information Test had five antecedants, a Technical Vocabulary Information Test developed by R N Hobbs and J W Thatcher (316 350–358), a Sports and Hobbies Participation Test devised by R R Blake and the writer (316 343–350), a Flying Information Test developed by the writer as a sub-test of the above (316.361), a Me-

chanical Information Test constructed by P C Davis and L. Hutchinson (316 323–327), and miscellaneous technical vocabulary items developed by F B Davis (316 361), all based on somewhat similar principles but focusing on different kinds of content, as the titles indicate The sports and hobbies test, for example, included items pertaining to driving a car, basketball, diving, hunting, building model planes, playing poker, motorcycling, and woodworking as active masculine avocations, and reading, music, etc , as sedentary feminine activities A sample item is

To "draw," a pool player hits the cue ball
- A at the right
- B at the left
- C high
- D low
- E don t know

These and other items were selected on the basis of several hypotheses 1) successful pilots, navigators, and bombardiers are differentiated by their personality traits and interests (e g , masculinity-femininity), 2) these traits manifest themselves in interest and participation in some activities and lack of interest and participation in others, 3) interest and participation result in the acquisition of specialized knowledge not acquired by others Particularly in the tests developed by or under the supervision of Blake and the writer, it was assumed that information which could be acquired only through participation, as opposed to observation, would differentiate most clearly the interested from the uninterested The activities or fields of knowledge tapped by the various information tests were selected on the basis of expected relationships between personality traits, interests, activities, and success in the three air crew jobs The items of each of the tests mentioned were selected first on the basis of internal consistency, then on the basis of validity (correlation with success in training) Only valid items were retained and incorporated in the General Information Test The validities of the antecedent tests for success in primary flying training (biserials r s with graduation-elimination) are given in Table 34, together with those for both the final form of the General Information Test for primary flying and, for an unselected experimental group, for both primary and all levels of flying training

The substantially higher validities for the experimental group can be explained at least partly by the unselected nature of the sample, for these

aviation cadets were sent to training regardless of scores on the various psychological tests used by the Air Force in order to obtain true indices of test validity The range of abilities being less restricted, the true relationships were revealed

It is interesting to contrast the approach of these information tests with that of the Michigan Vocabulary Profile Test While the latter used internal consistency as its criterion of item inclusion, and then proceeded tentatively to establish patterns of interest factor scores for various curricular groups, the Air Force information tests used factorial hypotheses as a basis for writing items, but included items in the scoring keys only as they proved to have individual validities for occupational prediction

TABLE 34

VALIDITY OF INFORMATION TESTS OF INTERESTS FOR PILOT TRAINING

Test	*N*	*r*	*Criterion*
Sports and Hobbies Participation Test	432–591	30 to 36	Primary School
Flying Information subtest	374–518	32 to 34	" "
Auto Driving subtest	371	39	" "
Hunting subtest	486	14	" "
Music subtest	118	− 18	" "
Reading (literature) subtest	287	− 14	" "
Technical Vocabulary Information Test	3151	17	" "
Mechanical Information Test	513–3151	23 to 32	" "
General Information Test	406–3146	17 to 21	" "
General Information Test (214 191)	1311	46	Experimental group Primary
General Information Test (214 191)	1311	51	Experimental group all schools

This was done in order to put valid tests into use at the earliest possible date The next step was a factor analysis of the tests to reveal what factors are measured and how unique they are, as Guilford has shown (316 817, 830–831), the information tests did measure a pilot-interest factor The next step would be to break this factor down by developing tests or subtests in which the items of the general information test are grouped according to hypotheses concerning the primary interest factors constituting the global pilot-interest factor, checking these for internal consistency and independence, making another factor analysis, and, if the tests seem promising, validating these purer factorial measures The first step in this last sequence was taken in the Flying Training Command at J C Flanagan's instigation (316 673–680) but was interrupted by the decline in training activities Work along these lines was resumed by F B Davis, J C Flanagan, and the writer (925 68–74) in the Personnel Distribution

Command in a study of combat leadership, the results of which were inconclusive but, insofar as they did reveal tendencies, showed that officers promoted most frequently while in combat tended to be more feminine than those who were promoted less often, while those who were promoted less often in combat seemed to be more masculine than the frequently promoted group This was exactly the opposite of the expected results, and the opposite of what inspection of significant items for the prediction of success in training had suggested. In training, it was the masculine, active, competitive, items on which the successful fliers did better than the failures As the training data were clearly significant and the combat data highly tentative, the latter relationships obviously need confirmation Also, the promotion criterion, although seemingly as good as any available, was shown in studies by J P Chaplin, H D Charner, W G Mollenkopf, and the writer (925 77–83) to be far from ideal as an index of success in combat flying

Although the wartime work of the Air Force with information tests of interest and personality factors was interrupted at the point described above, further work is being done both in and out of the services with these techniques They seem to the writer, who may be a biased observer in this instance, to be full of promise for the future

CHAPTER XIX

PERSONALITY, ATTITUDES, AND TEMPERAMENT

Nature and Development

THE field of personality is one of the most popular, challenging, important, and confused in contemporary psychology It was neglected by psychologists in the infancy of that science, studied by psychiatrists and psychoanalysts who used uncontrolled clinical methods, and then finally taken under consideration by psychologists who possessed scientific methods but too often lacked the orientation to persons as such which characterized the clinically trained medical men It is therefore small wonder that the psychology of personality has been in a chaotic state The origin and development of the theories of personality which one encounters today are hardly a topic for a book on the use of tests in vocational guidance and selection, treatments of the subject which were current when most of the available tests and inventories of personality were being developed will be found in psychological works by Allport (12), Brown (121), Shaffer (709), and Stagner (743) Murphy (554) has published a recent comprehensive treatment of the subject, which he also dealt with earlier in his collaborative synthesis of work in experimental social psychology (555) Hunt (391) has edited a generally excellent and up to-date symposium of encyclopedic dimensions and scope, the chapter on inventories is, however, unfortunately weak But it is relevant to consider the subject here from the point of view of the vocational counselor or personnel officer, from the perspective of the user of personality tests for vocational purposes

Definitions Some psychologists like to consider the personality as a whole, to think of it as a global unit, complex in nature but unanalyzable, a viewpoint often arrived at in the Gestaltist's protest against the unduly atomistic approaches of some Behaviorists To the scientifically minded person this point of view often seems mystical, vague, and of little value in practice Another approach defines personality in terms of the reactions

aroused in others, as social stimulus value To many psychologists this approach seems too limited in its empiricism, as it leaves the individual's personality in other persons, whose reactions are not completely uniform A third definition treats personality as a pattern of traits or ways of reacting to external stimuli Personality is then both analyzable and unitary, the operationalism of this definition appeals to the scientist The organismic or global approach to personality has something to contribute to this last viewpoint, for one can think of the individual as a more or less organized and integrated unit, and of the process of emotional development as one in which an attempt is made to organize a variety of reaction patterns or modes of behavior into an integrated, smoothly working whole One in whom a degree of integration appropriate to the demands made upon him by society has taken place is an emotionally adjusted person, while one in whom the integration has not taken place to the extent required by the demands of the environment, or one in whom the integration has partly broken down because of demands with which he was not able to cope, is an emotionally maladjusted or disturbed person

Psychologists interested in vocational guidance and personnel work seem to have found the concept of personality as a patterning of traits most helpful in their work, for discussions of emotional or personal adjustment and of personality traits abound in the literature and attempts to measure both general adjustment and specific traits and to ascertain their significance for vocational success have been numerous In an otherwise excellent discussion Warren (911 1015) states that the vocational counselor is less concerned with the degree of integration achieved by the client than with the nature and degree of his specific characteristics, for these determine his adjustments to his environment To the writer this seems to be too limited a view, for adjustment to the environment is partly a matter of adjustment to oneself, and adjustment to oneself is to a considerable extent a matter of the degree to which the various traits of one's personality are integrated In a well integrated personality the various internal needs and reactions to the various external pressures are harmonious the person is impelled, driven, or attracted in one general direction (minor needs and presses to the contrary being taken care of by the strongly integrated unit), and is therefore able to function effectively In the unintegrated or disintegrated personality, on the other hand, the reaction patterns are not harmonious, he is pulled and driven in various directions there is internal conflict, and functioning in society is im-

paired The vocational counselor and psychologist, and the personnel man who wants an effective employee, are therefore very much concerned with the degree and type of integration as well as with the specific traits which are organized into the whole

Role of Personality in Education and Occupation Approaches to the study of the significance of personality and temperament traits for success and satisfaction in school and at work have generally followed one of two patterns 1) the clinical, in which case-history material is cited in order to illustrate dynamics and document (if not prove) a theory, or, 2) the psychometric, in which reliance has of necessity been placed upon the imperfect instruments available for the measurement of personality In the former approach the findings prove little because of subjectivity and lack of controls, although they stimulate speculation, in the latter they prove little because of technical defects, although they do underline the need for better instruments The end result is that our current knowledge of the role of personality in education and in work is impressionistic or, when quantitative, superficial It has been shown by surveys of employment records, for example, that personality problems are the most common cause of discharge from employment (118,390) Case studies demonstrate that difficulties in learning to read are often caused by problems of parent-child relations, and observation led to the suggestion that some people considering engaging in social work are motivated by an unconscious desire to solve their own problems rather than to help solve those of others But none of these studies have yielded data which would enable one either to measure the extent and nature of the characteristics involved or to predict their interference or noninterference with success in any specific type of educational or vocational endeavor The conviction of their importance is strong and nearly universal, but the evidence is virtually lacking and the means of measuring the characteristics are sadly defective It is only for values and interests that techniques have been more adequate and results more conclusive, these have been discussed elsewhere

One reason for the lack of adequate objective evidence on the vocational and educational significance of personality traits is that students of vocational and educational adjustment have generally been specialists, not in personality, but in management, aptitudes, or instruction, while students of personality have generally been interested, not in vocations or in education, but in psychological theory or in clinical diagnosis Some of the personality inventories (e g , Bell, Bernreuter) are an exception to

this rule, but they suffer from the defects of the inventory technique, which are most serious in the field of personality, the more penetrating instruments (e g , Minnesota Multiphasic, Rorschach, Thematic Apperception Test) were devised for the study of personality organization or for the diagnosis of emotional disturbances. For our purposes, what is needed is a penetrating measure applicable and applied to occupational rather than to hospitalized populations

In view of the lack of sufficient objective evidence for a practically useful discussion of personality and vocational success, the results of what studies have been made will be reserved for the sections dealing with specific instruments Some comments are, however, called for in explanation of the failure to find clear-cut relationships between personality and occupations in the few studies which have been made with the more penetrating tests

Although it has been assumed that there should be linear correlations between certain personality traits and success in some occupations, for example social dominance and selling, submissiveness and bookkeeping, introversion and research or writing such relationships have in fact been found in very few occupations a somewhat higher degree of dominance has been found in salesmen than in clerical workers (587 201), but otherwise few significant differences have been reported The fact that some significant differences do exist and that some personality measures do have a degree of clinical validity, suggests that the general failure to find occupational personality patterns may be because personality is not related to occupational choice and success in the commonly expected manner Even in an occupation such as bookkeeping, a dominant individual may find outlets through advancement into supervisory and managerial positions, research may accommodate extroverts as well as introverts, for example, in sociological field studies, industrial chemistry, and the supervision of projects, and the literary extrovert may find outlets in public relations work, some forms of advertising and radio, or even fiction writing when formulas rather than creative imagination and insight are required A lawyer may be a bookworm or a dramatist, a scholar or a promoter a carpenter can work in morose silence, or exchange remarks and jibes with associates and passers-by between blows of his hammer, a packer may daydream or talk about the movies and the neighbors while placing batteries in cartons Roe's stimulating exploratory studies seem to confirm this hypothesis for artists (635) but to contradict it for paleontologists (636)

But if personality traits and temperament are not generally related to occupational choice or success, how, if at all, do they play a part in vocations? If the hypothetical examples given above are indeed valid, then personality as defined in this discussion determines the kinds of adjustment problems which the worker will encounter in any occupation he enters If he is outgoing and his associates withdrawn he will have one kind of difficulty, but it may be solved by changing associates rather than changing occupations, if he likes sedentary mental work rather than active contact work he may be a writer of books on his research rather than a promoter of the financing of more research or the administrator of a research project, if he is socially dominant the assembly worker may be the social leader or the thorn in the flesh of his fellows, rather than a follower or isolate in the group They will all be happy or unhappy in their work, depending upon the ease with which they make the modifications which it requires in their modes of behavior That such modifications are indeed made has been demonstrated not only with nursery school children by Page (583) and Jack (395), but also with college students by McLaughlin (498), although these studies did not demonstrate that the underlying traits were modified, they did show that the surface modes of adjustment were changed in ways which made the persons concerned function more effectively in their social groups Since personality traits have been defined as modes of behavior, they may be said to have been modified

If one were to ask, then, why bother to measure personality and temperament traits in personnel and vocational guidance work, there are two answers First, a poorly integrated personality (poor general adjustment) may have trouble adjusting in any training or work situation, and should either be screened out or given professional assistance in solving his emotional problems Second, a person with traits which are likely to make for adjustment difficulties in certain types of positions may be placed in a situation which is so structured as to turn his liabilities into assets or at least to minimize the chances of difficulty, he may be given psychotherapy to modify his personality in such a way as to facilitate adjustment, or environmental methods may be used to develop new modes of behavior which are more effective Many instances of maladjustment which appear at first to be vocational prove, after more careful examination, to be deep-rooted in the personality (257,442). When this is true, treatment by changing work situations or by on-the-job counseling may be necessary The reason for making a personality diagnosis in

vocational guidance and personnel work is, then, to screen problem cases and to assist in the making of more effective adjustments

Measures of Personality

Until about 1935 only two types of instruments for measuring personality and temperament traits were widely used in the United States rating scales and inventories These were both first put into extensive use and popularized during World War I, when Woodworth developed his Personal Data Sheet and various Army rating scales were experimented with, the details have frequently been written up, and will be found in Symonds (810) By 1935 several hundred personality inventories had been developed, but very few of them had been systematically studied after their first tentative launching, and the sophisticated segment of the test-using public had become wary of them (794) Rating scales also had proved disappointingly unreliable and invalid, but like personality inventories they were still used in many places either because the users were not fully aware of their limitations or, more often perhaps, because there seemed to be nothing better to use

In the thirties, however, another type of personality measure was introduced to the United States with the development of interest in the Rorschach Psychodiagnostik (644), a series of inkblots first devised as a projective technique by a Swiss psychiatrist by that name, and with the publication by Murray (557) of the Thematic Apperception Test, a series of semistructured pictures concerning which the subject makes up stories In these as in other projective techniques the examinee is presented with an ill-defined situation (inkblot, clouds, collection of toys, clay, or ambiguous pictures) and permitted to make what he will of it, the tendency is to structure it according to his own needs, thus revealing his personality traits unbeknownst to himself The clinician must then draw upon his own skill and insight to tease out the meaning of the figures, objects, scenes, or stories constructed by the examinee Although methods have been devised for obtaining seemingly quantitative scores from some of these tests, they are still essentially clinical techniques, rather than tests The fact that they appear to be more penetrating than personality inventories and have captured the interest of clinicians and researchers suggests that they will in time be greatly improved and transformed into more objectively scorable tests, but for the time at least they are limited to clinical use During World War II interest was revived in other little-used projective techniques, one adapted from a type of intelligence test

item the incomplete sentences test and the unstructured situation test These are still in experimental stages

In selecting specific tests for discussion in this chapter choices are limited to two types of instruments, personality inventories and projective tests, neither of which is presently very satisfactory or valuable to the vocational counselor or personnel man, and only one of which is of much value to the vocational psychologist Rating scales are not discussed, as they are filled out by persons other than the examinee and are dealt with in other texts (538,768) Space is devoted to inventories and projective tests, however, for two reasons 1) increasing use is being made of both types of measures in both personnel work and vocational counseling, despite widespread disillusionment with one type and skepticism regarding the other, and, 2) workers in the field need to know what has been done and is being done in the field of personality measurement, so that they may handle inquiries and take advantage of progress as it is made Personality tests and inventories are intriguing, it is well for the potential user to know the nature of their limitations in some detail

Two personality inventories are dealt with in some detail one, the *Bernreuter Personality Inventory,* because there is more published evidence concerning it than concerning any other inventory and because it is typical of many, the other, the *Minnesota Multiphasic Personality Inventory,* because it represents a somewhat different approach and has come into wide use and popular favor among psychologists Other inventories discussed more briefly are the widely used *Bell Adjustment Inventory* and the carefully constructed but new and less studied *Minnesota Personality Scale* One personality inventory developed for use in the wartime Army Air Force and no longer usable, the *Satisfaction Test,* is briefly described because of its implications for work with inventories in selection programs Many other inventories might be commented on, but the discussion of the above-named instruments should help the reader to examine critically the sweeping claims often made by publishers and authors

Two graphic projective techniques are treated in some detail, from the vocational counseling and selection point of view these are the *Rorschach Inkblots* and the *Murray Thematic Apperception Test,* both because of the widespread interest in them and because they are now being used in occupational research Two series of projective situation tests are described much more briefly, because of their possible significance for future work the series used by the Office of Strategic Services, and one

experimented with in the *Clinical Techniques Project* of the Army Air Force Finally, work with the *Incomplete Sentences Technique* is briefly discussed for the same reason

The Bernreuter Personality Inventory (Stanford University Press, 1931)

This personality inventory was based on earlier work done by Woodworth, Thurstone, Laird, and Allport, its principal contribution being its success in combining the items from several personality scales in one blank Although a commonplace today, this was then a time- and materials-saving novelty, and probably did more than anything else to give the inventory its widespread popularity When the studies published prior to 1941 were reviewed (794), the aggregate totaled 135, many more have since been published As the objective in this discussion is relevance to vocational guidance and personnel work rather than completeness no count has been made for the succeeding years those utilized alone amount to 27 The inventory is clearly still popular and widely used, despite a great deal of criticism

Applicability The Bernreuter Personality Inventory was designed for use with adolescents and adults Nothing has been found in the literature or encountered in counseling practice which suggests that the vocabulary and experiences sampled are inappropriate to those age levels

Age does not affect scores in relatively homogeneous populations such as those studied by Bernreuter (87), Carter (143), and Miles (528), although in more heterogeneous groups self-sufficiency and dominance seem to increase with age

It has been demonstrated that experiences planned to modify personality traits affect some types of scores on the Bernreuter (646,882) In the latter study the results may have been vitiated by training in the significance of behavior such as that described in the items of the inventory, for the experience was a course in applied psychology, Hartmann's findings (347) support this interpretation In the former the findings are more convincing, for the experience consisted of speech training provided for experimental but not for control groups, and the numbers were large

The effect of rapport has been investigated in a number of studies using the Bernreuter, the nature of the findings depending, as might be expected, upon the design of the experiment and the phrasing of directions Bernreuter (86) administered the inventory to students under normal conditions, then readministered it with instructions to answer it a) "as you would like to be," and b) "as you think you ought to be "

He found no significant differences, from which fact he concluded that the desire for social approval does not appreciably affect scores When somewhat different directions have been used, however, distortion of scores has been shown Olson (575) quotes an unpublished paper by Hendrickson which demonstrated that teachers retested with instructions to answer as though applying for a job made significantly more stable, dominant, extroverted, and self-sufficient scores than when answering normally, Ruch (656) found that college students raised their average extroversion percentile from the 50th to the 98th when asked to fake extroversion on a retest, and Fosberg (270) found that subjects instructed to make a good and then a bad impression on second and third testings succeeded in influencing their scores in the desired directions As the instructions in the last three experiments are more appropriate for testing the effect of conscious desire to fake than were Bernreuter's, it may be concluded that the desire to make a good impression, when it exists, does affect scores Bernreuter's directions and results do seem to warrant the conclusion that there is little if any disparity between the responses of persons in a nonevaluative situation (e g , students who know they will be marked on the basis of their achievement rather than on their personality inventory scores) and their responses when asked to reply in terms of their ideal selves, this only proves that the self-concepts of students differ little from their self-ideals

Mood might also be expected to affect scores on a self-descriptive scale such as the Bernreuter, but only two studies bear on that question One is Johnson's (406) comparison of the scores of 15 college women tested in periods of mild elation and again in periods of mild depression, in which only some differences approached significance, low moods being accompanied by slight shifts toward neuroticism, dependence, and submissiveness Johnson attributed the lack of significant differences to the freezing of responses once given to the Bernreuter items The case of a suicide was reported by Farnsworth and Ferguson (247), his neuroticism score changed from the 50th percentile 15 months, to the 83rd three months, before suicide Although the findings are by no means conclusive, the indications are that normal mood changes have no great effect on Bernreuter scores, while abnormal do

Content The Personality Inventory consists of 125 questions based on those used in earlier inventories, such as "Are your feelings easily hurt?" Answers are recorded on the blank, in terms of "yes," "no," and "?" There are few extreme or potentially offensive items, making the

inventory acceptable to most groups, with groups of adolescents, however, it is desirable to minimize opportunities for laughter and joking by businesslike administration and good proctoring

Administration and Scoring The inventory is self-administering, with no set time limit, and takes from 20 to 30 minutes Examinees sometimes ask what is meant by a question, for definitions of terms such as "frequently", although on the face of them these questions may seem warranted, the examiner must be careful to explain only unfamiliar terms, and to leave the interpretation of others to the examinee, as therein lies part of the significance of the test It is not so much the facts which matter in a personality inventory designed for the normal range of personalities, as the subject's attitude toward those facts, to make this concrete, it is not the actual number of times he has fainted that matters, as his feeling that he is or is not given to fainting

Scoring stencils are provided for neuroticism (B1-N), self-sufficiency (B2–5), introversion (B3 I), dominance (B4-D), self-confidence (F1-C), and solitariness (F2–5), with weights ranging from 7 to −7 assigned to each item according to its diagnostic value These weights were determined by relationship to the parent inventories Various brief scoring methods have been devised (18)

Norms These are provided on high school, college, and adult populations, gradations which are sufficiently refined as shown by studies of age differences The adequacy of the norms has been shown by several investigators (576,587,742,761) although some working with special populations have disagreed (948)

Standardization and Initial Validation Many of the items in the Personality Inventory were taken from the earlier blanks on which it was patterned, criterion groups selected on the basis of high and low scores on these other forms were then tested with the Bernreuter, and weights were assigned accordingly The correlations of Bernreuter's scales with the originals ranged from 67 to 91, as might be expected in view of the method of development This proved little concerning the validity of the inventory, as it depended upon the validity of the not-well-validated parent forms, but it did demonstrate what Bernreuter set out to prove, that one personality inventory could do the work of four It remained for subsequent studies, which Bernreuter himself failed to make, to establish the validity or invalidity of the instrument by the use of external criteria

Reliability The reliability studies have been numerous and are sum-

marized elsewhere (794), it need only be stated here that they have generally been found to be above 70 and often above 80, except after the lapse of substantial periods of time Whether the changes in scores which take place with time are the result of defects in the inventory or of changes in the subjects is not known

Validity The validity of a personality inventory for use in vocational guidance and personnel selection or evaluation must be considered from two points of view first, its value in screening maladjusted individuals who need psychotherapy or who should be rejected as employment applicants, and, second, its usefulness in predicting success and satisfaction in training and in various types of work Basic to this second purpose is another purpose, that of measuring identifiable traits which may be related to success and satisfaction They have also a third possible purpose, namely to assist in diagnosing the nature of a maladjustment, but that is one which concerns psychotherapists, rather than vocational counselors and personnel men The material discussed below has been selected and discussed with these distinctions in mind

The items in the Bernreuter were chosen on *a priori* grounds, on the basis, that is, of their diagnostic significance as seen in clinical experience They were validated by internal consistency, and named on the basis of an examination of their nature, thus one of his scales seemed to Bernreuter to measure autistic thinking, introspection, and other types of behavior warranting the name introversion (87) This procedure was criticized by Landis (452) as unsound because not empirical, he and Katz found that, although three-fourths of the self-descriptive responses of psychiatrically diagnosed neurotics agreed with objectively determined facts (451), some items are answered contrary to expectation (452) More normals than abnormals in their sample reported daydreaming tendencies, ideas running through their heads, etc The empirical approach was recommended, with items weighted on the basis of group differences (as in Strong's Blank) rather than *a priori* grounds

But Landis and Katz failed to take into account the important fact that Bernreuter had empirical evidence to justify his item weights, in the form of internal consistency data They therefore made no attempt to rationalize their findings with his, although both must be accepted This can be done by referring to the nature of the populations worked with Bernreuter's groups were college students, high- and low-scoring normals, while Landis' and Katz' were normals on the one hand and abnormals (neurotics and psychotics) on the other In other words, two

different types of "abnormals" were used, one in touch with reality, the other somewhat out of touch It is to be expected that the responses of these two groups would differ, for their degrees of contact with reality and their defense mechanisms are by no means the same Poorly adjusted normals may admit daydreaming more than well-adjusted normals, even though abnormal persons admit such behavior less often—they may not recognize it as daydreaming Two different sets of scoring keys may therefore be needed, one for more or less normal persons, and another for more seriously disturbed persons Bernreuter's scales were developed for use with normal subjects

The screening of maladjusted persons with this inventory has been studied by a number of authors, whose findings for psychotics and neurotics have been summarized as follows "When the data are examined in detail, they do appear to reveal differences between normal and various groups of abnormal individuals, even though these differences are not so clear-cut as one would wish unfavorable scores do tend to have significance, although favorable scores are not necessarily a sign of good adjustment" (794 100) Since the above summary two other studies have been published with military subjects Schmidt and Billingslea (677) found that although only the social-dominance scale clearly differentiated 329 psychopathic and neurotic from 95 normal soldiers, the patterning of scores on the Bernreuter scales was 80 percent effective in differentiating them Page (582) found highly significant differences in the mean neuroticism scores of large groups of medically diagnosed psychoneurotic and normal soldiers at Camp Lee These findings are in accord with a general tendency for personality inventories to be more valid in wartime military situations than in civilian life, a phenomenon which needs more study but which may be due to the fact that maladjustment in the armed forces is in a sense rewarded by escape from danger, and adjustment is in the same sense punished by the threat of death, whereas in civilian life the rewards go to the well-adjusted

Certain other types of problem groups have sometimes not been so well differentiated by the Bernreuter unmarried mothers did not differ from controls (570), and problem children at Mooseheart made scores comparable to those of others (732) But prison inmates have been found more neurotic than normals (381,172), Hargan's (331) contrary findings suggesting that traits may differ with types of crime Students coming to a college clinic for psychological help have been found more neurotic than others (761,664), college cheaters were more neurotic and dependent

than others (132,163), and the unhappily married were found to be more neurotic than the happily married (407,85)

The recognition of potential leaders by socially desirable scores on the Bernreuter has been shown to be possible in a number of studies Students who earn part of their expenses in college have been found more self-sufficient and dominant than others (664,101), fraternity members more stable, dependent, and dominant (101) Campus leaders have generally been found more dominant, self-sufficient, and stable than other students (664,394,631)

Ratings have been related to Bernreuter scores in a number of studies, particularly with college students as subjects A review of early studies (794 109) shows that these generally agree moderately well, the modal r being about 30 Two more-recent studies (944) found no relationship, however, suggesting that little weight can be given to validity studies based on ratings Both self-ratings and the ratings of others presumably have validity of a type, for one represents the subject-as-seen-by-himself, the other the subject-as-seen-by-others, even though the two images may not resemble each other, they are both important in the clinical study of an individual

Objective tests of intelligence have generally been found unrelated to Bernreuter scores (794 106), but stability, extroversion, and self-sufficiency have been found related to persistence test scores (661), and introversion to Rorschach-tested introversiveness to the extent of 78, affective stability to emotional stability 52 (892), findings partially confirmed in factor analysis of the two tests (475)

Grades have been used as a criterion with which to correlate Bernreuter scores in a number of studies elsewhere summarized (794 109), the general trend being for the relationships to be practically nonexistent In only one of the eight studies published prior to 1941 was any relationship found, in it, Neel and Mathews (565) reported that high-achieving students of superior mental ability were more introverted, self-sufficient, and solitary than low-achieving students of the same mental level The more refined approach to the problem used in this study seems to justify Stagner's statement (742) that personality affects scholastic achievement by influencing the use made of one's abilities and, therefore, does not yield a linear correlation with achievement More recent investigations have been published by Bennett and Gordon (71), and by Sartain (670), with nursing students as subjects, by Bryan (123) with art school students, and by Zelman (953) with general college students The first investigation

found little relationship, as determined by critical ratios, between nursing supervisors' ratings and Bernreuter scores, but Sartain reported correlations between grades and self-sufficiency, and between grades and social dominance, of .29 and .26. He discarded these as being of little value, and with only 81 cases the correlations are not reliable, but they do suggest that the inventory might contribute something unique to predictions normally based only upon intelligence and achievement. The other two studies showed insignificant relationships, confirming the common findings when intellectually heterogeneous groups are used. It seems clear that, if personality inventories are to be used in educational guidance, it should only be for the study of special groups such as underachievers.

Success on the job has been correlated with Bernreuter scores in relatively few studies, most of them fairly recent. Thirty foremen and assistant foremen were tested with the Bernreuter, Bennett Mechanical Comprehension, and Strong tests by Schultz and Barnabas (682), their criterion being combined ratings of budget control efficiency and employee relations. The combined scores of the tests had a validity of .52, the Bernreuter validity was .36 (only the one unspecified scale was used). A somewhat similar group of 40 foremen in an aircraft factory was tested by Sartain (671), again with ratings as the criterion. These had an interform reliability of .79, the validity of the predictors ranging from .01 (self-confidence) to .12 (social dominance), all of them too low to be reliable. Similar data for a group of 85 foremen yielded no better results, when 53 other foremen were classified as "good" or "poor" the difference in Bernreuter scores as not significant. Empirical keys for pilots developed in a study of aviation cadets initiated by the writer (316, 588–589) had no validity for success in pilot training.

Retail grocers, 70 in number, were rated according to credit and pecuniary strength on the basis of Dun and Bradstreet data by Hampton (327), these criteria yielded no correlations with Bernreuter scores exceeding .16. In personal contact saleswork as exemplified by casualty insurance salesmen, however, the relationships were higher. Bills and Ward (92) and Schultz (689) found that successful salesmen made more normal scores than did failing salesmen. Personality traits as measured by the Bernreuter therefore seem, like interests, to affect vocational success when the congeniality of the work is of especial importance.

Practice teaching has frequently been selected as an activity in which success might be expected to be related to personality traits as measured

by the Bernreuter Cahoon (130), Sandiford (665), Laycock (455), and Ward and Kirk (908) found no relationships using correlation techniques or critical ratios, but when Laycock compared the top- and bottom-quartile success groups there were great differences on all Bernreuter scales This finding was confirmed for another group by Palmer (585), and Pintner (605) found that good student-examiners (Stanford-Binet) were more stable according to the Bernreuter than were poor individual testers

In the one study of teachers in regular job situations, Gotham (302) failed to find any relationship between Bernreuter scores and teacher success, but the criterion was so unique as to need further study itself The subjects were teachers in 72 rural schools, their success being judged by "pupil gains" or the improved performance of their pupils In view of the many variables affecting learning, and the varying situations in which the teachers worked, the significance of pupil gains needs more detailed scrutiny than can be given to it here

A group of bank clerks were tested by McMurry (499), who found slight negative correlations ($-.27$ to $-.05$ for three different groups) between neurotic tendencies and efficiency ratings, these scores adding so little to the predictive value of the Otis that the relationship seemed unimportant

It may perhaps be concluded, from the above studies, that personality traits as measured by the Bernreuter are not generally related to success on the job, except in activities such as outside sales work in which the congeniality of the activity has a very direct effect on the degree of the worker s application

Success in obtaining employment or in retaining a job was not related to Bernreuter scores in the studies of the Minnesota Employment Stabilization Research Institute (587), but Morton (545), Christensen (156), and Lazarsfeld and Gaudet (157) have with both adults and adolescents found differences between employed and unemployed, job-getters and the unplaced, which were significant The employed tend to be more stable, more self-sufficient, and more dominant according to the Bernreuter

Occupational differences in scores on this inventory were first studied in the Minnesota Employment Stabilization Research Institute, where social dominance tended to distinguish salespeople from workers in skilled, semiskilled, and unskilled occupations, and policemen tended to be more dominant, stable, and extroverted than others, but other ex-

pected differences were not found in a cross-section of employed workers These trends were in general confirmed by Dodge (201,202) in New York City and Morton (545) in Montreal, both adding a few occupational differences salespeople were somewhat more dominant than clerical workers, traveling salesmen than bookkeepers (201,202), accountants and salesmen were most dominant, self-sufficient, and stable, engineers and unskilled workers least so, while professional men and executives tended to be dominant, carpenters and electricians tended to be emotionally stable (545) Motion-picture writers studied by Metfessel (526) did not differ in individual traits from the general population, but examination of their average profiles showed that the patterning of their scores is not typical Johnson (403) also found 150 salesmen to be dominant on the Bernreuter when compared to the norm groups, they were a homogeneous group in this respect, unlike an equally large group of seminary students, McCarthy (492) also studied seminarians, finding that they tended to be somewhat unstable and self-conscious when compared to the general population But in all of these instances the overlapping of groups was so great as to make application of the findings impractical

Stability in an occupation and *job satisfaction* should have been common subjects of study by means of personality inventories such as the Bernreuter, as it is commonly assumed and case studies have shown (257) that personal maladjustment often underlies vocational dissatisfaction and frequent job changes Only one such study has been located with this inventory, however, in it, Seagoe (689) found no significant differences between teachers who stayed in that occupation and those who left it, although there was a tendency for the well-adjusted, and for the maladjusted of lower intelligence, to remain in teaching, and for the maladjusted of superior mental ability to give up teaching, as though they had the insight and the ability to leave an uncomfortable situation More studies of this type seem desirable, to throw more light on the dynamics of vocational adjustment

Use of the Bernreuter Personality Inventory in Counseling and Selection There is some danger, in a summary of this sort, that the discussions of group differences which justify statements such as "salesmen tend to be more dominant than clerical workers" will leave in the mind of the reader the impression that a person making a high dominance score on the Bernreuter might be a salesman, and that conversely a person making a low score would do well to avoid sales work It is well to remind the reader that the existence of group trends is compatible with the finding of many individual exceptions some salesmen are not dominant, and

some dominant persons would not make successful salesmen When it is remembered that social dominance is just one characteristic often found in good salesmen (in life insurance they tend also to be over 35, married, fathers, to have bank accounts, and to carry insurance themselves) the reason is obvious It is well to remember this in reading the following summary

B1-N, the emotional instability or neuroticism scale, appears to measure emotional sensitivity Low scores tend to indicate a wholesome extroversion, an ability to face facts and the environment objectively and to deal with them without internal conflict, whereas high scores suggest unwholesome introversion, poor adjustment to the environment and a tendency to withdraw from it A great variety of maladjusted people make high scores on this scale neurotics, autistic schizophrenes, and depressed persons Low scores are made by emotionally stable people, and by those who in different situations are aggressive rather than withdrawing, and also by leaders, fraternity members, the happily married, the paranoid, manic individuals, and hyperthyroids It has some occupational significance, as shown in the tendency of the employed to be more stable than the unemployed (this could be either cause or effect), the superior stability of policemen, accountants, salesmen, carpenters, and electricians, and the tendency of emotionally stable teachers to stay in their field while the able-unstable changed to other occupations

B2-S, the self-sufficiency scale, probably measures another type of introversion The high-scoring person tends to be self-sufficient, does not depend on others for advice and emotional support, he is not withdrawn so much as free from the necessity to advance, an introvert in the Jungian sense of the term The low-scoring person is probably not an extrovert, however, in the usual sense, for this implies a wholesome turning to the environment whereas in such instances the turning outward is the result of a need to depend upon the environment for emotional support normally found within the self Low scores therefore probably represent an unhealthy sort of extroversion, contrasted with the wholesome extroversion measured by B1-N Maladjusted groups which tend to make high self-sufficiency scores include neurotics (the false self-sufficiency of compensatory fantasy?), withdrawing persons (for the same reason?), and divorcees, those making low scores include cribbers and epileptics The occupational significance of this scale is indicated by the high scores made by leaders and contact workers, and the low scores made by those who work primarily with records or materials

B3-1 has been found to resemble B1-N to such a high degree (794 110)

as to justify not using it The introversion-extroversion which it was designed to measure has, we have just seen, already been provided for

B_4-D, the dominance-submissiveness scale, measures the tendency to dominate in face-to-face situations It is apparently not a pure trait, but a combination of wholesome extroversion and sociability (794 110) Low scores indicate submissiveness, but high scores may be indications of the conviction that one should seem dominant rather than of a tendency to take the initiative in social situations Problem individuals who tend to make high scores seem to include only those who react aggressively to difficult situations (794 116), if these may indeed be called problem people, low scores tend to be made by withdrawing persons and by others who have difficulty coping with the environment (794 116) The occupational significance of the scale is shown by the superior dominance of the employed, salespeople, policemen, accountants, professional men, and executives, and the submissiveness of the unemployed, unskilled and semiskilled workers, clerical workers, and bookkeepers

The F1-C and F2 S scales, developed by Flanagan on the basis of factor analysis (261), have been respectively shown to have much the same significance as B1-N and B2-S

In schools and colleges the Bernreuter can be used with a fair degree of confidence as a measure of group trends, and for the screening of problem individuals who are to be studied by more intensive methods It is likely to prove more helpful in survey testing than as part of a battery for intensive study of an individual "Bad" scores, which are sometimes high and sometimes low, can generally be assumed to have some significance, but "good" scores may be compensatory rather than the result of a wholesome adjustment It should be more useful in educational institutions than in clinics or employment situations, because other methods of study suitable for clinical use should prove more penetrating in mental hygiene work, and because the desire to make a good impression can distort scores when applying for employment The item-validity is such as to make the inventory best for use with normals and near-normals, rather than with psychotics The inventory is of questionable value in detecting behavior-problem cases, as opposed to otherwise emotionally maladjusted persons The use of the Bernreuter scores in counseling concerning vocational choice appears to be virtually limited to consideration of the significance of dominance scores for business contact occupations When these are high, confirmation should be sought in extracurricular and leisure-time activities, when low, case history and cumulative record ma-

terial should also be examined in order to ascertain how the trait has affected social behavior, for some successful salesmen are not exactly dominant individuals, although these men may perhaps expend more energy in meeting the social demands of sales work than do more dominant persons In any case in which abnormally high or low scores are made, the counselor should study cumulative record and interview data in order to understand the significance of the score for the person in question, and if it indicates that the counselee may have difficulty making adjustments which he is likely to be called upon to make, the counselor should make an attempt to get him the needed therapeutic help

In guidance centers the use of this inventory may be similar to that in schools, especially if used routinely for survey testing and screening. With clients who come because they themselves feel the need for help it may provide one more kind of data concerning personality traits, to be viewed in relation to other data, but it is not likely to help with those who come because they are sent or who are referred for appraisal as possible employees It has proved of some vaue in selecting salesmen, and so may have a place in personnel evaluation, whether because it portrays the applicant's actual personality or because it shows how well he knows what a good salesman should be like and do, by and large, however, other methods of personality study should be relied upon when referral or recommendation for employment is under consideration

In employment services and in business and industry inventories such as this are not likely to prove satisfactory, because of their transparency, except as pointed out in the preceding paragraph The occupational differences which have been observed with it were all detected, it should be remembered, in situations in which the examinee had little or nothing at stake

The Minnesota Multiphasic Personality Inventory (University of Minnesota Press, 1943, Psychological Corporation 1945)

This personality inventory was developed by Hathaway and McKinley at the University of Minnesota as a clinical instrument for use in psychiatric diagnosis (353) It was not intended as a test for use in educational and vocational counseling, or in personnel selection Their purpose was to develop one personality inventory which would measure all aspects of personality which bear on psychiatric diagnosis, thus implementing Rosanoff's theory of temperament (645) They wished to make more objective the judgments that are reached in a clinical situation by providing

more systematic coverage of behavior and attitude items than is generally possible in an interview But there is evidence in many guidance centers of an interest in applying this instrument to vocational guidance and selection, apparently in the belief that since it is considered a better clinical inventory than most others on the market, it should also be a better vocational test This is a nonsequitur, but it makes desirable some consideration of the test in this chapter No attempt will be made to go into its clinical validity in any detail, as that is a long story the tangential relevance of which makes a mere summary suffice, what is known of its vocational significance will be discussed at somewhat greater length

Applicability The Multiphasic was designed for use in mental hygiene and psychiatric clinics, with older adolescents and adults who have had a few years or more of education It has been administered to junior high school boys and girls, but according to the authors (352 1020) it has not been validated at that age, for which many items might conceivably have quite different significance The authors report that several of the traits measured change within relatively short periods of time, as one would expect of depression and hypomania, which are attitudinal manifestations of one underlying type of temperament Some of the other traits might be expected to be less subject to the effects of experience and of mood, as in the case of masculinity and psychopathic deviation Studies of the extent and nature of such fluctuations have apparently not been published

Content The MMPI consists of 550 self-descriptive items such as are found in the Bernreuter and in other personality inventories on which it was based They are classified under 26 categories, ranging from general health through the gastrointestinal system, habits, family, occupation, sex, phobias, and morale to items designed to show whether the examinee is trying to describe himself in improbably good terms Some items such as the first two listed below are quite innocuous, while others, like the last four, are more likely to seem offensive

I like to read newspaper editorials
I hate to have to rush when working
Someone has it in for me
Peculiar odors come to me at times
At times I feel like smashing things
There is something wrong with my mind

Administration and Scoring There is no time limit, but testing normally takes from 30 to 90 minutes, depending on the education and ad-

justment of the examinee There are two forms of the test, one consisting of a set of cards administered individually and to be sorted into three stacks (True, False, Cannot Say), the other a booklet with IBM answer sheets The test authors recommend the individual form, and Ellis (238 423) has suggested that it may be superior to the booklet form, but Wiener's study of 200 veterans in a guidance center found no differences in group trends on the two forms (924)

The card test is recorded on special forms, and both are scored by means of stencils, or the booklet can be machine scored Scoring may now be done for nine reaction patterns hypochondriasis, depression, hysteria, psychopathic deviation, masculinity-femininity, paranoia, psychasthenia, schizophrenia, and hypomania Others may be added Four other scores (question, lie, validity, and a "suppressor variable") are also available to aid in judging the meaning of the scores It should be noted that although at least one of the traits may be thought of as one aspect of temperament (masculinity-femininity), two others seem to be mood-manifestations of another aspect of temperament (hypomania-depression), and still another may be the pathological extreme of a personality trait (schizophrenia), the others are traits made up of modes of behavior which are not normally considered as components of the normal personality, but are generally thought of as clinical syndromes or even disease entities On logical grounds one might therefore question the soundness of applying such measures to normally adjusted persons and drawing conclusions concerning occupational differences, but to do so is consistent at least with Rosanoff's theory of temperament (645) This theory postulates three components, of which the above psychotic tendencies are developments and on which this inventory is based

Norms The standardization group consisted of about 700 men and women representatives of the general Minnesota population in age and education, and not under medical care, norms are based on hospitalized patients from each of the nine diagnostic categories, averaging about 50 in number (manual) The development of norms for psychiatric classifications is difficult because of the impurity of cases in actual practice, and the consequent difficulty of classification in any one category, clinical users of a test such as this should examine published data on the norm groups more carefully than is appropriate here No occupational norms have been published, but data are given for five small groups of workers in as many occupations in two published studies (469,893) discussed below

Standardization and Initial Validation In attempting to develop a measure of Rosanoff's temperament components Hathaway and McKinley relied partly on the only existing inventory which had the same purpose, the Humm-Wadsworth Temperament Scale The data concerning this scale have been so uniformly favorable when published by the scale's authors or persons working under their auspices (365,387,388) and so often unfavorable when analyzed by others (228,316,902), and its handling has been a matter of such frequent criticism, that ethical and informed psychologists are reluctant to use it despite some good and unique features They also drew from the Bernreuter and the Bell, which were used in their first investigation (353), and made up other items of their own on the basis of psychiatric manuals and clinical experience Items were assigned to scales on the basis of the extent to which they differentiated 221 classified psychiatric patients from 724 normal persons bringing relatives or friends to the University of Minnesota Hospital, 265 college-entrance applicants at the University, and other similar persons presumed to be normal The first clinical group consisted of 50 carefully screened hypochondriacs (496), a cross validation group of 25 hypochondriacs, and control groups of 699 normals, 50 normals with physical disease, and 45 miscellaneous psychiatric cases The hypochondriacs were significantly (C R = 10 9) distinguished from the normals, the other non-normal groups were also, but the overlapping in their cases was much greater (C R = 4 0 and 2 5) The other clinical groups were equally small the depressed also number 50 (354) But the tendency to distinguish appropriate groups from others was in each case cross-validated and stood the test The various scales of the Multiphasic may therefore be said to have been empirically developed and validated against appropriate external criteria

Reliability The test authors believe that the nature of the Multiphasic precludes the possibility of adequate indices of reliability (352 1020), because of the variations of some of the traits from time to time within a given person, and because of the heterogeneity of the items which make up clinical syndromes in contrast with pure traits They have reported, however, that the test-retest reliabilities range from 71 to 83, this is about as high as those of most personality inventories An empirical check on the authors' hypotheses concerning variations in scores would be desirable, it would be possible, for example, to rate clinically studied individuals on these traits, and to relate changes in rated condition to changes in inventoried condition, thus ascertaining whether the somewhat lower

than desirable reliabilities are due to variations in the individual rather than to the unreliability of the instrument Although such ratings are themselves not very reliable, if made each time by the same person they would presumably have a sufficiently high degree of reliability as indices of increase or decrease in the type of behavior under study

Validity The *clinical* validity of the Multiphasic was reviewed by Ellis (238) in 1946, by which time 13 clinical validation studies had been published Ellis' approach to personality inventories was hypercritical he defined r's of 0 to 19 as negative, 20 to 39 as mainly negative, 40 to 69 as questionably positive, 70 to 79 as mainly positive, and 80 up as positive, although he claimed on page 393 to "evaluate the reported coefficients of correlation in terms of the conventional estimations" a claim subsequently modified Nevertheless, he found that eight of these investigations yielded positive results, while three showed some validity and only two failed to demonstrate validity in this inventory (the writer's summation from data on pages 420 to 422) These figures were much better than those for the other inventories summarized by Ellis, the next best being only nine confirmations of the Bernreuter's validity, while six studies showed some validity and 14 showed none, according to Ellis' severe and unconventional criteria These data suggest that the Minnesota Multiphasic has more validity for screening and classifying personality problems than any of the generally available personality inventories Findings of some of the specific studies are discussed below, but no attempt is made to review them all in detail as clinical diagnosis is not the central interest of this chapter

In the development of the scales Hathaway and McKinley found that, despite overlapping of populations, from 50 to 80 percent of each their psychiatrically diagnosed groups were differentiated from normal persons and generally even from each other by the scales for hysteria, hypomania, psychopathic deviation (355), hypochondriasis (496), psychasthenia (497), and depression (351) Although their groups were small, ranging from about 25 to 50 per category, the clinical diagnoses were carefully made and the trends were very suggestive They were confirmed by most subsequent studies, as brought out below

The inventory was administered to 85 naval psychiatric patients by Benton (73), who found that five out of ten schizophrenics were differentiated by the appropriate scale, as were five out of nine hystericals, 13 of 16 delinquents (psychopathic deviation), and nine of ten homosexuals (femininity) In another study he and Probst (74) tested 70 persons diag-

nosed by Navy psychiatrists, in this the psychopathic deviate, paranoid, and schizophrenic scales showed statistically significant differences between clinical groups and normals, although the differences for the other scales were not clearly significant

Delinquent adolescent girls were compared with nondelinquent controls by Capwell (137), who found that the former were clearly differentiated by all but the hysterical scale, the psychopathic deviate scale being the most diagnostic As Van Vorst (891) reported negative results with a group of psychopathic delinquents, the subject needs further investigation

Other psychiatric groups were studied by Gouch (304), who found differences between the scores of normals and 136 neuropsychiatric soldiers classified according to severity of neurosis, or as psychopathic deviates and psychotics, Harris and Christiansen (343), who tested 53 psychiatrically diagnosed patients and found perfect agreement in more than half the cases, and complete disagreement in about 10 percent of the cases, Leverenz (468) who used the test in an Army hospital and found it of "definite value" despite some disagreement with clinical diagnoses, Michael and Buhler (527), who tested 90 psychiatric patients in a general hospital and found it successful in only about 45 percent of their cases and of little value in differentiating psychopaths from psychotics, and Schmidt (676), who found statistically significant differences between normal soldiers and those diagnosed as constitutional psychopaths, neurotics, and psychotics

Ellis suggests three possible explanations of the positive results generally obtained with the Minnesota Multiphasic as opposed to the more commonly negative results reported in clinical studies of other inventories (238 423)

1 Individual administration may bring about, at least in part, the same kind of rapport factors which are so important in case study interviews

2 Most individual administrations have been done with one test, the Minnesota Multiphasic, which was standardized on a decidedly clinical and objective, rather than the more usual subjective (internal consistency), basis and which, in consequence, may possibly be a superior questionnaire

3 The majority of Multiphasic validity studies have either been done on groups (similar to those) used to standardize the test, which have been institutionalized populations which may be more sophisticated and more honest than other abnormal groups, or else they have been done with military personnel, who may have every incentive to answer personality questionnaires honestly

Obviously, further investigations of these hypotheses are needed, and will probably be reported in the literature in due course Wiener (924) has already shown that, at least with veterans coming to a guidance center, the means of those taking one form are the same as those of the men taking the other form If this is verified by other approaches to the problem of form, in which the group form is given to groups and the individual form to individuals (procedure apparently not followed by Wiener, who used the two forms under identical conditions), Ellis' first hypothesis must be discarded

Achievement criteria of various types are of principal interest to vocational guidance and personnel workers, who need to know not only the effectiveness of the test in screening maladjusted persons who may need special attention, but also the significance, if any, of the trends measured for educational and vocational success No studies of the educational predictive value of the Multiphasic have been noted in the literature, but one paper advocating the use of the inventory in vocational counseling, one study of the relationship between Multiphasic items and occupational success, and four on occupational differences revealed by the test, have been located These are discussed below

The "accumulated experience" of two veterans' counselors who used this inventory in vocational counseling was described early in 1945 by Harmon and Wiener (339) As the test was less than two years old at the time at which they wrote, work with veterans was only getting under way, and their data were not quantitatively treated, their statements should probably be viewed as hypotheses to be investigated rather than as findings to be applied in practical work Statements such as one to the effect that the Multiphasic "has proved an instrument of prime utility" which "has served to delineate personality characteristics of crucial importance in the actual choice of a vocation and has yielded valuable information to aid in prognosis of success in training" were made, it should be noted, *before* the veterans in question had tested the choices made in actual work and *before* they had had any opportunity to achieve success or failure in training The case studies presented are more convincing as evidence of the usefulness of the inventory in locating persons who need psychotherapy before they can function in any kind of work, than as evidence of its value in aiding in the choice of occupation or of type of training in only one of the six cases did it really play a differential role in vocational counseling

Success in flying training was the criterion used in a study reported by

Guilford (316 599–601) The group form was administered to 856 would-be pilot cadets in 1944, and the items were validated after reports of their success or failure in primary training became available It was decided to validate the published scales only if a sufficient number of items were valid to justify scale validation The *phi* coefficients were unimodally distributed with a central tendency at zero, indicating that few if any items had any genuine validity for success in flying training The clinical scales were therefore not correlated with the criterion

Occupational or preoccupational differences were studied in two investigations by Lough (485,486) In the first she found that 185 unmarried women undergraduate students of education were a relatively stable group with a very slight tendency toward hypomania and that there were no significant differences between those preparing to be elementary or music teachers In the second paper she reported findings for 300 unmarried women undergraduates, including the original group and 115 liberal arts college students A slight tendency toward hypomania was found in the new group as in the original, suggesting that this might be characteristic of adolescents There were no differences between curricular groups, to which nurses and the various liberal arts majors were added She concluded "it is not a useful instrument for differentiating between those who are more suited for one occupation than another The primary value of the MMPI seems to be to give some insight into the emotional life of the individual and to detect those who may be in need of psychological or psychiatric counseling" It should be noted that her first conclusion is based not on success but simply on choice (a somewhat questionable criterion, as some who choose fail and some who might succeed do not choose), and that her second conclusion is based on the evidence of other studies, reviewed in earlier paragraphs of this section The writer is inclined to subscribe to her conclusions, but the first one at least needs further proof

Women clerical workers, department store saleswomen, and women optical workers were tested by Verniaud (893), the samples numbering 40, 27, and 30 respectively The workers came from several different offices and stores, and from several departments of one factory The profiles of the two white-collar occupational groups differed very little from the norms, except for somewhat low hypochondriasis in the clerical workers and decidedly masculine scores in the salesclerks, but the optical workers were decidedly hypomanic and psychasthenic, and somewhat paranoid and psychopathically inclined

In view of what is known of the interest patterns of women clerical workers, it is not surprising to find them a normal group, resembling women in general The masculinity of department-store salesclerks is surprising, as their work is a relatively passive type of sales and they deal largely with feminine items, as Verniaud points out, this finding would bear further investigation

The findings concerning the factory women raise several questions, for they may be peculiar to the local situation (one company in one town), to the occupation (blocking, roughing, emery grinding, polishing, finishing jobs), to the socio-occupational level, or to the population (e g , a minority group) There is no description of the status of the women but the factory workers were all employed on war jobs (Navy contracts), whereas the others were engaged in more normal, peacetime, operations rather than in war industries This suggests that they may have been a quite atypical group of women workers drifters, thrill seekers, and others who might flock to a boom industry on a temporary basis Verniaud does not go into this possibility, but does state that "In terms of the expected meanings of the characteristics (MMPI scales), we would expect these workers as a group to be restless, 'full of plans,' alternating between enthusiasm and over-productivity in energy output and moods of depression, more inclined toward anxieties and compulsive behavior than the average individual, disinclined or unable to concentrate for long periods on one task, somewhat oversensitive or suspicious of the good-will of others, somewhat more inclined than the average woman to disregard social mores' Only three sample case studies are presented in this report of her master s thesis, but Verniaud states that the test profiles are borne out by case-study material which she collected in the thesis Before any vocational guidance or selection applications are made of such findings, it would be imperative to ascertain whether the factory workers whom she studied are in fact typical of women factory workers in general, this type of occupation only, only this plant at this time, or merely of women war-plant workers This last type of group no longer has occupational significance but must, if accurately described by these findings, still be unhappy and making unhappiness for others

Life insurance salesmen and women social workers, 50 subjects in each group, were studied with the Multiphasic by Lewis (469), each group being compared with the norm group of the same sex The insurance salesmen were found to be significantly more depressive, hysterical, psychopathic, feminine, paranoid, and hypomanic, the last-named had the

highest T-score (58 1), while only femininity and hysteria reached a T-score of 55 The social workers were significantly high on the depression and hysteria scales, significantly low on those for masculinity, hypochondriasis, psychasthenia, and schizophrenia, their psychopathological sophistication being perhaps a contributing factor to their low scores, for this reason pre-training tests, evaluated after employment in the field, would have provided more convincing evidence of occupational differentiation in this type of work Lewis also found that those whose interests, as measured by the Kuder-Preference Record, were least appropriate for their work, tended in each occupation to be the least well adjusted, but the differences were not clearly significant in most comparisons

Job satisfaction has not been studied by means of this inventory, although the findings just reviewed have implications for that topic if confirmed by other studies

Use of the Minnesota Multiphasic Personality Inventory in Counseling and Selection In a few years there will probably be enough accumulated evidence concerning the traits measured by the Minnesota Multiphasic to justify a discussion of their significance paralleling that for the Bernreuter, but all that can be written at this stage of its development would concern their clinical significance rather than their vocational implications Such material has a very important place in a manual of clinical psychometrics, but not in a book designed for use in counseling concerning vocational choice, selection, or upgrading Not, that is, until more is known about the vocational significance of clinical data It is enough for our purposes to state that the authors' claim that persons who make extreme scores on any of the scales probably need psychotherapy seems valid, as high scores have generally been found to characterize appropriate clinical groups A high score may be defined as a T-score exceeding 70

Occupations which may be appropriate or inappropriate for those who make extreme scores on this inventory cannot as yet be listed, if indeed they ever will be We have seen that one investigator concluded that the real value of the instrument is in clinical rather than in vocational diagnosis. There are indications that hypomania, hysteria, and femininity may be characteristics which make for success and satisfaction in selling life insurance, that depressive and hysterical tendencies may be suggestive of social work for women, and masculinity may make sales work a suitable outlet (other things being equal) for women Other possible vocational implications of this inventory need confirmation with larger

and more representative groups whose background and working environments must be carefully described for the data to be meaningful In the meantime Vernaud's suggestion that the Multiphasic be used only as a clinical instrument seems to this writer to be the only one justifiable at this stage of its development

In schools and colleges the Minnesota Multiphasic Personality Inventory may therefore be useful as a device for screening students in need of further study and perhaps of counseling in relation to personality adjustment, more often, it is helpful as a diagnostic device following such screening by other less elaborate inventories or after referral by other staff members, to provide the counselor with some orientation to the nature and extent of the maladjustment It is not recommended as an aid to vocational counseling except when the counselor is also a clinical psychologist and the client is a maladjusted person in need of help with an immediate problem of vocational choice or adjustment

In guidance and employment centers the Multiphasic has more of a place because the larger number of persons with personality problems who come to such centers makes careful screening imperative This inventory may therefore be helpful in a secondary test battery when a shorter routinely administered personality inventory, the psychometrist's or preliminary interviewer's observations, or the referring source suggests the presence of psychopathology Positive findings would then be an indication of need for therapy beyond the scope of the typical vocational or placement counselor, or for co-operative work with a psychotherapist, the vocational counselor helping the client to make a vocational adjustment which contributes to his general adjustment by making one aspect of his life that much more successful and satisfying Differential occupational prediction on the basis of Multiphasic scores, such as is suggested and practiced by some counselors, is still premature except in a highly tentative way and on the basis of confirmation by case-history material In evaluating persons being considered for referral for employment or referred for evaluation by employers the inventory may have some value as a screening or selective-placement device, but in view of what is known about the faking of scores on other inventories the results in such cases should be very critically viewed

In business and industry this inventory may be helpful as a means of screening out maladjusted employment applicants, as those who make high scores are extremely likely to have personality problems, but low scorers may include many who are merely successful as disguising their

true characteristics It may also be of use in personnel evaluation, either for the selective placement of handicapped persons or for the improvement of supervisory and executive functioning In this type of work the interpretation should be done only by a qualified clinical psychologist, as the results might otherwise be bad both for the individual and for the company, and referral facilities should be available if psychotherapy is indicated The inventory may prove to have value in the selection of salesmen and other contact personnel, and perhaps with other types of employees, but local validation and normative studies must be carried out before such use is possible

The Bell Adjustment Inventory (Stanford University Press, 1934, 1937)

This is another widely used personality inventory, particularly in schools and colleges, but it has not had particular appeal either for clinicians or for industrial personnel workers Published three years after the Bernreuter and scorable for four aspects of adjustment, somewhat different in superficial ways from the Bernreuter and its predecessors, it somehow escaped the more violent criticisms leveled at them and caught the second crest of the wave of popularity of personality inventories Perhaps it seemed sufficiently different from others to be "worth trying" as the search for an effective personality inventory continued New types of inventories such as the Minnesota Multiphasic Personality Inventory had not been published as yet, the Humm-Wadsworth was criticized by the Bernreuter's critics and projective techniques had not yet become generally known Bell's monograph (62) gave users of his inventory the feeling that they knew something about the instrument, and the names of the traits it measured had a safe and homely sound quite different from the trait-names of the much criticized inventories

Description The Bell Adjustment Inventory is published in two forms, one for students and one for adults, and is scorable for four aspects of student and five of adult adjustment home, health, social, emotional, and, in adults, occupational adjustment It is designed for use in high schools and colleges, and with adults Although it has been suggested that some of the items make it offensive to some people, Pallister and Pierce (584) reported that they found it quite acceptable to the Scottish subjects with whom they worked The blank consists of about 100 questions like those in other inventories, although some of the questions which are treated as health questions by Bell are weighted for neuroticism in most inventories (e g , "Do you have many headaches?")

Responses are of the yes-no type, and are recorded on the blank or on IBM sheets It is self-administering, with no time limit, and requires no more than 30 minutes Scoring is quickly done by means of stencils, each response being given a weight in one scale only, and the score is the sum of the circled responses Norms are provided for high school students, college students, and adults, these have been criticized by Tyler (886) because of the use of a five-point scale which gave undue weight to changes of a few responses, Bell (63 995) considers the norms tentative despite the lapse of more than ten years since publication, and recommends the development of local norms The reliability of the inventory has generally been found satisfactory for group purposes but somewhat low for individual diagnosis, Turney and Fee (884) reporting retest reliabilities ranging from 74 to 85, and Traxler (857) odd-even reliabilities of from 83 to 93

Validity In the development of the forms (62) items were used which distinguished the high- from the low-scoring groups of students or adults on whom they were standardized Scores were correlated with those obtained by existing inventories, and the coefficients ranged from 57 to 89 for appropriate scales Students and adults designated by counselors who knew them as well or poorly adjusted in each area were found to be distinguished quite significantly by appropriate scales

The clinical validity of the Bell has been disappointing Ellis' summary of published studies (238) reports 12 investigations of the inventory, 11 of which showed that it had little or no value for the identification of maladjusted persons, and only one of which showed positive results Readers who wish to look into the details are referred to Ellis' concise and well-organized, even though severe, summary The writer has located only one investigation missed by Ellis, and although it (938) is favorable one such study cannot change the picture presented by studies such as those by Marsh (511) and Feder and Mallett (252) in which the inventory was found to have very little value in screening students in need of psychotherapy

Grades were correlated with Bell scores by Drought (213), Young et al (950), Clark and Smith (160), Crider (181), and Griffiths (313) with results which were negative, they generally have been in such studies Only Fischer (256) has reported positive results with this inventory, using an index different from those of the other studies He constructed an underachievement ratio based on scholastic aptitude and grades, which correlated 42 with Bell's emotional adjustment score, suggesting as in some

of the studies with other personality inventories that when intellectual factors are taken into account, personality traits play an observable part The defect may therefore lie, not so much in the inventory, as in the design of the studies Fischer found a correlation between emotional adjustment and point-hour ratio of − 32, as his cases numbered only 48 his findings are merely suggestive, but may be worth following up

Success on the job has been correlated with Bell scores in only one known study, in which Forlano and Kirkpatrick (268) tested 20 women radio tube mounters No data are given for the Bell alone, but only for a combination of its social adjustment score with that of Washburne's inventory Twenty cases are too few for the results to be conclusive, but it is interesting that all eight employees rated "good" in efficiency by their supervisors made average or better scores on the inventories, while the 12 who were rated "fair" made average or below average adjustment scores

Occupational differences have not been studied in employed persons by means of this inventory, but McCarthy (492) administered it to seminarians, and found them below average in total adjustment on the Bell as on the Bernreuter Whether this was a personality pattern which existed before entrance into the priesthood, or merely a transitory reflection of the experiences these men were undergoing in training, was not brought out by the study

Job satisfaction has not been studied with this inventory, although the inclusion of a number of questions bearing on this in the adult form might have been expected to be the result of or to encourage such studies Only Seagoe (689) has touched on this subject in her study of permanence in teaching, in which, as we have seen, there was a slight tendency for well-adjusted student-teachers to remain in the profession, together with the less-intelligent maladjusted, while the brighter maladjusted tended to leave for other types of employment, but these differences were not statistically significant

Use of the Bell Adjustment Inventory in Counseling and Selection Unlike other personality inventories, this instrument attempts to measure not only traits (emotional adjustment or stability) but also degrees of adjustment in several areas (home, social groups, and health) This seems to have been done on the assumption that it would be helpful to know which area is the most active source of maladjustment, which the source of most security and satisfaction The intercorrelations of the several keys reported by Bell (62), Tyler (886) and others (r's = 04 to 53) indicate that there is some overlapping of the scales, but they are low

enough to suggest the conclusion that several factors are being measured But the Bell total adjustment score has a correlation of .77 with Bernreuter's emotional stability scale (602), and Turney's criticism that, "It must require considerable faith or temerity to believe that 140 items, averaging 35 to a division of the kind in the scale (the four scales), really measure adjustment in a satisfactory manner The complexity of the psychobiological environment must have been grossly overestimated by a host of psychologists if we are mistaken about this" (884), seems to contain a possible explanation of the low intercorrelations It may be that each set of 35 items is merely a sample of the items which would make up an emotional stability scale, their low intercorrelations being due to the inadequacy of their sampling of the various ways in which emotional stability or neuroticism manifests itself This has been suggested also by Young, Drought, and Bergstresser (950) If this is so, the study of foci of maladjustment may still be helpful, but the danger of making the deduction that a certain individual is "well adjusted socially but not emotionally" should be clearly recognized

The occupational significance of the Bell is unknown, as no adequate investigations have been made

In schools and colleges this inventory may have some value as a screening instrument for the location of maladjusted students, but other measures have been proved more effective The value of the part scores as indices of the foci of maladjustment has hardly been demonstrated, and the writer believes that once problem cases have been located by other screening instruments such diagnostic matters can be better handled by interviews or by projective tests such as the Thematic Apperception Test There is no evidence that the inventory has any value for directional vocational or educational guidance

In guidance centers also the use of this instrument hardly seems warranted by what is known about it Other inventories can screen more effectively and diagnose more significantly, and the clinical techniques are in any case rather readily used in such situations

In employment services, business, and industry other inventories and tests which have been studied with vocational purposes in mind have been demonstrated to have some value, whereas there are no data which indicate that this measure will help in personnel work

The Minnesota Personality Scale (Psychological Corporation, 1941)

This inventory is included in this chapter, not because there is any evidence of its validity in vocational counseling or personnel work, nor

because it is widely used and needs to be understood, but simply because it is the most recent step in the evolution of a number of attitude and personality scales which have been carefully studied They have contributed to our knowledge of social psychology if not actually to our proficiency in vocational psychology The first milestone was the development of a technique for the study of attitudes by Thurstone (841), simplified and applied more specifically to the study of morale by Likert (471) and used effectively in a study of attitudes and unemployment by Hall (324) The technique was further refined in an intensive psychometric study of the effects of the economic depression of 1929–39 on personality, carried out by Rundquist and Sletto (658) and resulting in the Minnesota Scale for the Survey of Opinions (the intensity applied more to the psychometrics than to personality) This inventory gave scores for morale, feelings of inferiority, family attitudes, attitudes toward the legal system, economic conservatism, attitudes toward education, and general adjustment, attitude variables which it was thought might be affected by prolonged unemployment The present scale is the result of a factor analysis of this and several other inventories by Darley and McNamara (192)

Description This inventory consists of five parts, or a total of 218 questions, the sections being designed to measure morale, social adjustment, family relations, emotionality, and economic conservatism Typical items are

Court decisions are almost always just
There is really no point in living
Do you have a fairly good time at parties?
Do you and your parents live in different worlds, as far as you are concerned?

The answers are arranged on a five-point scale of frequency or intensity, depending on the trait It is designed for use in the last two years of high school, in college, and with adults, there are two forms, for men and for women It can be administered in less than 45 minutes to persons of these educational levels, and is scored by stencils or by IBM machine Norms are for 2000 men and women freshmen at the University of Minnesota local norms would be needed if the inventory were much used, because of the differences found in the attitudes measured by some of these scales with differences in economic status and degree of sophistication (379) The scales are quite reliable, ranging from 84 to 97 when computed on a corrected odd-even basis (manual)

Validity The items were selected after a factor analysis and other studies of the Minnesota Scale for the Survey of Opinions, the Bell Adjustment Inventory, and the two Minnesota Inventories of Social Attitudes, which showed that the 13 scores of these inventories could be explained by the five factors measured by the scales now comprising the Minnesota Personality Scale The inventory is therefore internally consistent, and incorporates the best elements of its predecessors Thus refined, the Bell items take on a different character, for they are part of internally consistent and relatively distinct factors for example, the best Bell health items became part of the emotionality scale The authors believe (manual) that these new scales should have at least the validity of the parent scales, although this does not suggest much value for the Bell-derived scales, it should be remembered that they were technically improved and perhaps given new validity (which still needs to be proved), and the Minnesota Scale for the Survey of Opinions was shown by Rundquist and Sletto (658) to have value for the study of the attitudes and adjustments of the unemployed Validation of this instrument against external criteria is needed before it can be useful in practice

Ratings of corresponding traits in 235 student nurses were made by their supervisors and colleagues in a study by Bennett and Gordon (71), which showed little relationship between the two sets of data This has generally been the case when ratings have been correlated with inventory scores, and may only prove that ratings are of little value

Success in flying training was the criterion used in a study initiated by the writer and completed by Guilford (316 601–603) A group of 338 would-be pilot cadets who took the test early in 1941 were sent to pilot training, and subsequent reports of their success and failure were correlated with the scale's five scores The biserial coefficients of correlation ranged from − 09 to 04, showing no validity for this purpose No other validation studies have been located

Use of the Minnesota Personality Scale in Counseling and Selection As there is no objective evidence on which to base suggestions for the use of this attitude and personality inventory in educational and industrial personnel work these paragraphs are limited to a few suggestions concerning possible values Even the best predecessors of this inventory were never more than attitude-research inventories which were not put to use in personnel work to any appreciable extent Despite this, the Personality Scale is technically good enough to merit research in practical situations Were it not for a few items which would probably not be acceptable to

many workers (e g , one on the C I O), it might be considered for use in morale surveys, when it is desired to obtain data not only on satisfaction with the specific aspects of the job and working conditions usually covered in such surveys, but also on general morale and aspects of emotional adjustment which are often the underlying causes of job dissatisfaction The acceptability of the items to employees must first be ascertained, if unacceptable in their present grouped form, some could be modified and they might be made more palatable by putting them in omnibus form and thus burying the more personal items in fairly innocuous material Similarly a counselor in an educational institution who desires data concerning the climate of student opinion may find a survey with this scale helpful, it would not need modificaion for college use Those whose scores deviate considerably from the mean may be observed or interviewed in order to ascertain the effect of their atypicality on their status in the group This application might also be made in industrial situations if respondents were identifiable, but it seems likely that best results would be obtained by securing anonymous responses when administrative action might be feared

The Rorschach Inkblot Test (Grune and Stratton, first published in Switzerland in 1921)

This series of inkblots was developed by a Swiss psychiatrist, Hermann Rorschach, as a measure of the underlying structure of the personality, was experimented with by him and his students for a number of years before it was introduced into the United States during the 1930's, and has grown rapidly in popularity as a clinical instrument since that time, becoming practically a cult in some circles The result has been a vast amount of publication concerning it, and some research, most of the writing being concerned with its use in personality study and clinical diagnosis Although some proponents of the technique have advocated its use in vocational guidance and selection (e g , 607), little evidence has been adduced to justify or contradict the sweeping claims made for it

The fundamental differences between this and other types of tests, the varied aspects of the personality which it is purported to measure, the internally consistent logic upon which it is based, and the dramatic use to which it has sometimes been put, have given the Rorschach a wide appeal, at the same time, the enthusiasm of its proponents and the extent to which it has been based on clinical intuition and subjectively rather than quantitatively analyzed experience have antagonized many more

scientifically minded psychologists However, the proper approach is an open-minded study of the instrument in which one can assess its demonstrated value and establish hypotheses concerning its potential value, which can then be tested experimentally It is in that spirit that the writer has attempted to deal with it in the following pages, for though he has used the inkblots both in research and in clinical practice, and believes them to be of value, he is not a "Rorschacher" or a cultist

To attempt to treat the clinical validity of this complex and subjectively scored test is unfortunately too sizable a task for a book such as this To explain the technique alone requires a whole volume, as has been shown by Rorschach (644) and later by Beck (55,56,57), Bochner and Halpern (108), and Klopfer and Kelly (433), a volume on its validity is also needed, but has yet to be produced The pattern followed in discussing personality inventories will therefore be departed from, and this section will attempt only to describe the test in sufficient detail to provide an orientation to the procedure and to the nature of the test, and to discuss the studies which have been made of its significance for educational and vocational counseling and selection Its clinical validity will not be treated, a decision which seems justified also by the fact that the inkblots are in any case a diagnostic rather than a screening device

Description The Rorschach Inkblot Test was designed originally for use in the diagnosis of psychiatric disorders in adults It has since been used, however, with normal adults, adolescents, and children, and has been found applicable to any person of school age provided the interpretation is made in terms of the age group to which the examinee belongs The test consists of ten white cards, on each of which is reproduced one large inkblot Some of the inkblots are monotones (gray), while others include color The test is administered individually in clinical and sometimes in personnel practice, the examinee telling what he thinks each inkblot might be, the examiner recording responses on a blank which includes outlines of the inkblots, this is followed by an inquiry in which further details concerning responses are elicited, and by a testing of the limits, in which the psychologist ascertains whether or not the examinee is capable of giving certain types of responses which he has not previously given (56,433) When used for screening (as it occasionally is), for personnel selection, or for research the test is often administered as a group test, the inkblots being projected onto a screen and the examinees recording their responses on diagramed blanks, this is followed by a modified inquiry, in which the examinees locate their

responses for the examiner, there is no testing of the limits (346) A multiple choice form of the group test has also been developed (346), of doubtful value as seen below

Scoring the Rorschach is a time-consuming task when it is desired to obtain a detailed clinical picture of the person being studied, and often takes two or three hours When Munroe's inspection technique (551) is used merely in order to derive an index of total adjustment the time may be reduced to 15 minutes per examinee In either case the person doing the scoring must have had intensive training in the use of the test, combined with a good background in clinical psychology, for despite the lengthy and helpful discussions of scoring now available (57,433) the procedures are quite subjective Some users of the test are, in fact convinced that to objectify the procedure would be to destroy its clinical value (433 20–21, 432, 57 vii)

The norming of the Rorschach has also been a sore point with many psychologists In general, Rorschachers have felt that clinical experience and insight are sufficient to justify the interpretations commonly made, and Rorschach's original insights are often appealed to as evidence of the significance of a response Others have been concerned with the accumulation of norms for the various types of responses, for various normal and clinical groups, in order that the clinical significance of a response might be objectively demonstrated and verifiable by reference to quantitative data For example, Beck's first monograph on the inkblots was quite normative in its approach (55), but his two later books (56,57) have been more subjective and more dependent on clinical intuition It is this very lack of objective norms for many aspects of the test which makes clinical training and experience necessary to the users of the Rorschach, it also makes essential a scientific attitude and a tendency to seek objective evidence to justify clinical intuitions The problem is not as simple a one as to collect or not to collect norms, however, as the scoring and interpretating often require the relating of one variable to others in ways which do not lend themselves well to quantitative treatment as we now understand it

As responses and scoring have so far been discussed in abstract terms, it may be well to make the subject tangible by describing some types of responses and their scoring One inkblot, for example, may look to the examinee like a leopard skin, and the inquiry may explain that this is because of the shape and because of differences in the furry texture In scoring this response three items are of interest the examinee responded

to the *whole* picture rather than to details, seeing the inkblot as a unit rather than as a number of disparate units, he responded thus partly because of the *form,* and partly because of the *texture* which he used as color These three items are added to similar items obtained in response to other pictures, giving scores, respectively, for the *W* response, *F,* and *Fc* Interpretation of the test then proceeds on the basis of each of these scores, seen in the light of other related scores *W* is thought of as revealing a tendency to respond to wholes, to organize and synthesize, a high score is taken as revealing superior intelligence, unless it is so high or so superficial as to take on another meaning *F* is thought of as a sign of emotional control, although if it is high and certain other indices are low it may mean rigidity *Fc,* or the use of texture, is interpreted as a sort of shock absorber, of controlled sensitivity to the environment A variety of other modes of response to the inkblots, and the content of the responses, are also analyzed, various ratios are computed, and a profile is plotted in order to facilitate study of the pattern of responses A verbal summary or personality sketch is then prepared on the basis of this analysis Most of the justification for these interpretations, it should be emphasized again, lies in the intuition of clinicians who have used the test and studied the responses of persons whom they had come to know well by other clinical methods Only a few of them have been validated by objective methods

Validity It should be clear from what has preceded that the validity of the Rorschach in personality diagnosis has been demonstrated largely by the extent to which clinicians have thought it agreed with psychiatric diagnoses Studies with the Rorschach have been reviewed by Hertz (367,368), subsequent reviews have been published by White (921) and Kaback (414) The studies reviewed below are selected because they deal with the vocational significance of Rorschach indices

Grades in college were used as a criterion against which to validate the total adjustment score of the Group Rorschach in a study by Munroe (552) Her subjects were students at Sarah Lawrence College, where grades were not those usually given for specific course work, but faculty ratings of academic standing, a more general evaluation of the student's status The correlation was 49, as contrasted with one of 39 for the A C E Psychological Examination It would be desirable to have similar data for colleges in which more traditional marking methods are used.

Success on the job has been studied with the Rorschach only in unpublished investigations, so far as this writer knows One large depart-

ment store has gathered data on executives employed during recent years, testing them during the selection process and relating their scores to ratings of subsequent success Although the group in question is still small (N = 30) and the results tentative, they appear to be promising, however, preliminary findings such as these are often reversed when studies are completed The Group Rorschach was used in a study of aviation cadets made with the assistance of the Josiah Macy Foundation early in World War II (oral report to officers of Psychological Research Unit No 1 by Miss Sadie Sender), the design of which was defective in that cadets tested *after* failure were contrasted with successful pilots other studies had shown that eliminated cadets showed many symptoms of depression early in the war, thus making their Rorschach scores questionable It was administered also to 660 aviation cadets tested in the Aviation Psychology Program of the Army Air Forces in 1943, low validities of doubtful significance being found for a few scattered single indices which, when combined, gave a biserial coefficient of correlation of 17 with success in pilot training (84), however, when this formula was cross-validated on another group of 156 cadets it had a negligible validity of 04 (797 555, 316 633) The Multiple-Choice Rorschach was validated against success in pilot training with negative results (316 636), in other similar studies the results were no better The only conclusion one can draw from these various studies is that if the Rorschach has validity for the selection of personnel for various types of work (or, by implication, the counseling of people concerning the appropriateness of vocational choices), there is as yet no evidence to indicate just what single or combined Rorschach traits might confirm one choice or contraindicate another

Occupational differences as shown by the Rorschach have been studied by Kaback (414) and by Roe (635,636) Kaback used the Group Rorschach results of 300 pharmacists and accountants, dividing them into professional and preprofessional (student) groups She found point-biserial coefficients of correlation (to be distinguished from biserial coefficients) of 54 and 65 between 24 Rorschach components and professional or preprofessional group membership, in other words, there was a statistically significant relationship between Rorschach pattern and occupational group membership Kaback points out, however, that the overlapping of groups is so great as to make the application of her findings to individuals highly questionable The picture is further confused by the finding of equally great differences between the employed and student groups

(point-biserials of 625 and 62), the thumbnail sketch of the employed pharmacists having practically no resemblance to that of the student-pharmacists although those of the two accounting groups are more similar The sketches of the two professional groups are summarized here as illustrations of Rorschach results

Pharmacists intelligent adults whose impulse control functions well in general with one limitation their conscious repression of impulses (F% = 47) plays a relatively great role, and inner stability a relatively smaller role (M = 2 09)' Fairly marked amount of anxiety (presence of K and k responses) which is counterbalanced by sensitivity to inner and outer conditions (FK + Fc FK + K = 2 28 1 19) Intellectual flexibility marked(W, D,d,S present) Spread of interests somewhat limited (H + A 5 other content categories) However, in terms of general adjustment, the group falls within the general range

Accountants superior adults Well-balanced impulse control, function smoothly in conscious impulse control (F% = 44), rational behavior in emotional situations (FC CF + C = 92 59), inner stability (number M present) This group has a tendency to attend more to stimulations from within than to external stimulations (M sum C = 3 1) and to use them productively (W M = 12 3) Conscious control refined by use of shock-absorbing functions (FK + F + FCR = 56) by being sensitive to inner and outer conditions Small amount of anxiety (some K and k) and a slight tendency to overcautiousness in emotional contact with outside world (FC + CF + C Fc + c + C' = 1 5 3) Good mental elasticity (W,D, d,Dd,S present) with widespread interests (H + A 8 other content categories) In general, a well-adjusted group

Artists were the subjects of Roe's first study (635) They were a group of 20 eminent American artists whose co-operation was obtained in an investigation of the effects of alcohol on the creative process Her results showed that the group was extremely heterogeneous on the Rorschach general adjustment scores ranged from 3 to 18 with a mean of 10 3, about that which Munroe found predictive of maladjustment in college (552) and higher than the mean of 7 7 which she found in paleontologists No control group was used, all other comparisons being with the extremely subjective standards of Rorschach tradition The artists tended to make more than an average number of whole responses, responded more to color and shading than men-in-general, and gave rather more than the normal proportion of anatomical and sexual responses, neither of them surprising in trained artists The protocols of the tests were submitted

to a noted Rorschach authority (Dr Bruno Klopfer) for "blind" analysis (interpretation without benefit of other case material), his only data being age, sex, and the fact of their being professionally successful Two of the protocols made it obvious that the men in question were connected with art Klopfer noted this, and stated that one was probably a successful creative artist, but that the other was so lacking in creative ability that it was improbable that he could be successful at it professionally Creative ability was noted in five others, but said to be limited in one and unusable in another because of neurotic conflicts, in five others it was said to be absent, and he implied its absence in three more, no relevant comments were made concerning the other five, which implies no notable creative ability These findings are important, for Rorschachers have without objective evidence set much store in the inkblots' ability to reveal creative ability, whereas these two competent Rorschachers (Klopfer and Roe, who made a similar analysis before she asked Klopfer to check hers) failed to find signs of creative ability in 15 out of 20 eminent artists As Roe points out, creativity may not be required for success in art in our culture, but the law of parsimony would seem to require one to question an unvalidated assumption concerning a test before questioning cultural standards Roe's general conclusion was that despite some trends, as noted above, "there is no personality pattern common to the group"

Vertebrate paleontologists and technicians assisting them were tested in the other study by Roe (636) The two groups numbered respectively 16 and 9, tested at their annual meeting with the Group Rorschach The general adjustment scores of the scientists averaged 7 7, of the technicians 9 4, not a significant difference and both below 10, which may tentatively be considered the critical score for maladjustment The three best-established scientists made an average adjustment score of 4 Unlike the artists, these two groups were found to be quite homogeneous in personality patterns Both groups tended to give whole responses, but as would be expected those of the scientists were superior in quality to those of the technicians, in keeping with their higher mental level Unlike the artists, the scientists as a group gave only one sex response (several later said they had consciously suppressed these), while the technicians gave more than the average number of human anatomical and sex responses The scientists gave an unusually large number of animal anatomy responses, as might be expected in a group of men whose work involves spending hour after hour with bones, the tech-

nicians gave fewer, perhaps reflecting less absorption in their work than their professional counterparts Roe classified such responses as "technical", because appropriate to the profession, that medical students give more anatomical responses than men-in-general (346) is further evidence of the effect of interest and experience on test scores The most striking finding was the very small percentage of human movement responses, considered indicative of creative imagination the group appears thus to have a decided tendency to react objectively to the outer world, to avoid projecting themselves into situations and structuring them in terms of their own needs It is also interesting to note that the three most successful men have what would be considered sufficient movement responses, that is, enough creative ability to rise to the top of their profession, in which the more completely objectively minded worker normally does well Color shock, or inability to handle color, which is considered indicative of inability to handle social relationships effectively, was also common in this group of men whose work permits them to live in relative isolation and carries few social obligations Roe related her findings to a study of Munroe's (553) with college girls, which suggests that the personality patterns and vocational relationships indicated here may exist before entry into occupations In view of this, her summary concerning this group of scientists takes on especial significance

> the men who follow this vocation show, as a group, certain definite characteristics of personality structure They tend to abstractions, to formalized, objective, thinking, with a marked inhibition of any tendencies to project themselves into a situation They empathize little, either with things or with other people, and they have a rather passive emotional adaptation There is further indication that within this group, those who have been able to maintain objectivity and at the same time not inhibit creativity, those who can to some extent at least project themselves, are the ones whose work is most broadly theoretical and most widely significant *Caution because of the small sample should be invoked here*, yet the indication is entirely logical (636 326)

The italics are the writer's, for the results of a highly subjective test, based on groups of 16 and 9 persons, with no adequate controls, can be considered no more than tentative But they are the most challenging of any study so far made of the relationship between personality and vocational choice, and indicate that the technique should be further developed and that other groups should be studied with it, in order to

add to our knowledge of vocational psychology and to the tools of vocational guidance and selection

Use of the Rorschach Inkblot Test in Counseling and Selection No attempt has been made to assess, in this section, the validity of the Rorschach as an instrument for the clinical study of personality, although it is obvious that such validity would be helpful in counseling and evaluation because of the insights it would give into the types of adjustment problems an individual might encounter and the amount of difficulty he might have in handling them The making of such an assessment would require more space than is warranted in anything other than a textbook of clinical psychometrics Attention has been limited to the relationship between Rorschach scores and vocational choice and success

The studies so far completed show that no relationship has been found between Rorschach indices and vocational success, although one study now in progress appears more likely to yield positive results Studies of occupational differences are by no means conclusive, in one instance because of excessive overlapping of groups despite significant differences and because of differences between employed and student groups, and in another because failure to find homogeneity in the occupation is questioned by the absence of a control group, in a third group the numbers are so small and controls so lacking as to necessitate drawing only the most tentative conclusions from what is otherwise a most revealing and challenging study

In view of the above, the Rorschach can be considered only an instrument which may be worth using in validation studies, as one which research may yet prove extremely valuable in vocational counseling and selection, but about which too little is now known to justify its use in practical personnel work

The Murray Thematic Apperception Test (Harvard University Press, 1935, 1943 Grune and Stratton, 1949)

This projective technique is, even more than the Rorschach, a clinical device rather than an objective test, and one the occupational significance of which is unknown It is briefly described here for two reasons it has so challenged the interest of test users that questions concerning it are common, and it has promise as a research technique for the study not only of personality adjustment but, more specifically, of the determinants of vocational choice and satisfaction Unlike the Rorschach,

it is not a measure of the structure or organization of personality, but rather a technique designed to bring out the content of the personality, the needs, strivings, and environmental pressures which are felt by the person being studied This fact might lead one to question its potential value as a device for use in directional vocational counseling or selection, were it not that the needs or strivings which it reveals may well be the determinants of vocational choice and vocational interest

Description The TAT, as this test is generally called, was designed for use with older adolescents and adults, but pictures have since been added which make it administrable to older children and younger adolescents the examiner merely selects the appropriate pictures As most of the studies made with it have been made with the older group, however, more is known about its scoring and interpretation at that level The test consists of a series of 20 pictures for a given age and sex group The pictures are semistructured, that is, their content is more like a specific object or scene than is the content of an inkblot or a cloud picture, but expressions are sufficiently ambiguous and action poorly enough defined so that it is possible for the subject to project himself into the situation and shape it somewhat according to his own needs and fears Thus one scene depicts a human figure seated or kneeling next to a seat, a small object on the floor or ground before him, head bent and face hidden To one person this figure represents a boy who has just broken his mother's favorite vase, at the remnants of which he is staring, to another, a girl who has just shot her lover and, dropping the pistol in front of her, is overwhelmed by her deed, to someone else it is a young man, fondly gazing at a flower given him that night by his sweetheart Each person sees what he needs or wants to see in such a picture

The test is administered individually, sometimes the examiner, and sometimes the examinee, writing down the examinee's story of how the scene came about, what is going on at the moment, what the characters feel, and what the end result will be Scoring methods vary with the objectives of the examiner, and might better be called interpretive methods, for they are neither objectively based nor objectively expressed Instead, the examiner analyzes the content in order to determine the underlying themes (hence the name of the test), to ascertain whether or not the plots are happy, logical, probable, and to find out with what kinds of heroes the subject identifies himself and the forces to which he feels subjected The manual presents a somewhat more quantitative but time-consuming method for obtaining a weighted count of the needs

(e g , abasement, aggression, dominance) and forces (e g , affiliation, aggression, loss) affecting the hero or examinee, a scheme useful when research is being conducted in group differences or in relationships between test and criteria The norms in this scoring system consist of the responses of normal college students, in certain other methods there are none, and the data are used simply as clinical or case-history material to be interpreted in the light of other personal data to make a dynamic and meaningful picture of an individual It is obvious that, like the Rorschach, this test can be used only by well-trained and experienced clinical psychologists Bellak (64) and Tompkins (850) have published scoring aids and manuals, each differing from the others in important aspects

Validity The most intensive clinical validation of the technique is reported by Murray (557) in a study of Harvard undergraduates, which showed a high degree of consistency between TAT and other clinical evaluations made independently Harrison (344) found that conclusions based on it agreed well with case history material and psychiatric diagnoses in a mental hospital As the question of clinical validity is not one of primary concern in this context, however, these and related investigations will not be gone into in any detail it is important only that there are some indications of validity in what is still a clinical device which seems likely to develop into a test

Occupational differences in TAT patterns have been touched upon by Roe in her study of the personalities of artists (635), the one published study in which the test has been applied to occupational groups (Neal E Miller, John L Wallen, and the writer used it with aviation cadets during World War II, for a clinical study of success and failure in flying training in which all data were merged to yield a dynamic picture of each cadet rather than to reveal group differences in test scores) Roe found the test difficult to administer to her 20 artists, as they were so critical of the artistic quality of the pictures that they found it difficult to focus on the telling of a story Interpretation of results was made correspondingly difficult The content of the stories was not unusual, a tendency toward feminine and nonaggressive identifications was noted, otherwise there is little that seems significant in the data, of which Roe does not seem to have pushed the analysis as well as she did that for the Rorschach

Use of the Thematic Apperception Test in Counseling and Selection This brief account of the TAT has attempted to make clear its embry-

onic status and at the same time to suggest its promise as a device for measuring, more subtly than any personality inventory, the needs which drive people and the forces which they feel pressing upon them Although virtually no use has been made of the instrument for vocational counseling or selection, and none should at present be made, the technique is one which should be developed to a point which will make it useful in studying the needs and drives which are related to vocational choice and success, and for ascertaining the relationship between these and satisfaction in various types of work It would be helpful, for example, to know that the need for winning affection is more often satisfied in social work or in teaching than in medicine or law, and to have an objective method of measuring that need Such developments in the TAT are remote, but they are mentioned in the hope that research will be prosecuted which will bring them about

Trends in the Measurement of Personality

Perhaps the major trend in the development of instruments for the measurement of personality during the past 20 years has been one away from the inventory technique and toward various projective devices, illustrated by the disfavor with which personality inventories are generally viewed and by the rapid growth in popularity of the two best-known but complexly scored projective tests At the same time there has been a minor trend of considerable importance, an interest in the refinement of inventorying techniques, illustrated by the publication of factor-analysis-based forms such as those by Guilford (318), Darley and McNamara (192), and others, and by the empirically based Minnesota Multiphasic (353)

Both trends can be traced to the low validity which has been seen generally to characterize personality inventories In the first it caused a search for a more subtle and penetrating type of test which would probe underneath the sophistications and rationalizations of the subject in order to get at the structure and content of his personality, in the second it resulted in a greater emphasis on purity of factors in some inventories and on empirical weighting in others The improved personality inventories seem to this writer to be better stopgaps while the subtler projective techniques are being objectified and validated

These trends have manifested themselves in other ways which have as yet made less impact on applied psychology, but which should be

familiar to the practicing vocational counselor or personnel worker, and which warrant experimentation by personnel psychologists These will be very briefly described in the following paragraphs

Custom-Built Personality Inventories A series of standard personality inventories, including the Bernreuter, Adams-Lepley, Humm-Wadsworth, Minnesota Multiphasic, Minnesota Personality, and the several Guilford scales, were administered to aviation cadets who later went to pilot-training schools during World War II As reported in various official bulletins and summarized by Guilford (316 Ch. 23), none of these had any validity for pilot selection

The Shipley Personal Inventory, developed for wartime use by the Office of Scientific Research and Development (563,564), also had no validity for pilot selection (316 604–607), although it did have validity for certain other types of military selection and for screening combat-fatigue patients (925 115–121) It is of interest because of its effective use of the forced-choice technique, in which the subject must choose between two sometimes innocuous but often offensive self-descriptive items

The Satisfaction Test (801,316 736–745) was one personality inventory which did have some validity for pilot selection It was developed by Robert R Blake, John L Wallen, Joseph Weitz, and the writer with the specific conditions of military life and wartime flying as their content, for example

> If given the choice and having equal opportunity and ability, would you rather
>
> 53 A ambush the enemy?
> B storm an enemy position?

The keys were empirically developed, on the basis of item validation against success in training They were twice validated and cross-validated on groups ranging in size from 800 to 2000 cadets, and each time had a low but statistically significant validity of about 20 This would not have been sufficient to justify using the test, had it not been that its extremely low correlation with the selection battery made even this validity a unique addition to the predictive value of the battery Conclusions drawn from this study have been summarized as follows (801 744)

1 When a valid battery of aptitude tests has been developed and new aptitude tests are found merely to measure the same thing in different

ways, thereby adding little to the validity of the existing battery, personality inventories may be worth considering

2 In such a situation, the personality inventory may have low validity, both absolutely and relatively to the aptitude tests, but, if the relationship to the criterion is significant, it will have a unique contribution to make to the battery

3 Standard personality inventories are less likely to be valid, because of their general terms and situations, than custom-built inventories based on analyses of the behavior and attitude-evoking situations in the vocation or in the employing organization

4 Empirically validated success-failure keys, checked against the logic of the situation and of the item, are likely to prove more valid than keys based on clinical judgment or on an internal rather than external index of validity

Situation Tests The situation test is one in which the examinee is put in a partly rearranged but real-life situation and his behavior noted and analyzed The technique was first developed by German psychologists (245), was experimented with in the selection of reserve officers at Harvard under Murray, and was used extensively by the Office of Strategic Services under Murray's direction during World War II (33,558) It was relied upon there, despite its cumbersomeness and lack of proved validity, because it was felt that the screening of superior men and women for confidential assignments in which effective social relations, leadership, and discretion were vital could not be better done in any other way The tests were administered during a sort of house party in which 18 candidates, seven psychologists, psychiatrists, and sociologists, and eight junior psychologists participated for three and one-half days A variety of standard tests, performance tests, projective tests, interviews, psychodrama, and casual observations were used, but the techniques of interest in this context are a series of leaderless group-situation and individual situation tests In the Wall Test, for example, six examinees were assigned the task of getting a heavy eight-foot log over two parallel ten-foot walls, set eight feet apart, without touching the ground between the walls this gave opportunities to observe leadership, social relations, initiative, practical problem-solving ability, etc In the Construction Test the candidate had to build a five-foot cube with a glorified tinker-toy, aided by two helpers whom he was to direct, the helpers were junior psychologists whose task it was to turn lazy, recalcitrant, and insulting in order to test the candidate's frustration tolerance, the task was never com-

pleted and some candidates became either very upset or enraged by their humiliations

Observation during these situation tests made it possible to rate candidates for emotional stability, social relations, energy and zest, leadership, security, and other traits, and a staff conference synthesized the findings into a job-fitness rating and an evaluation note The criteria of success used were far from perfect, but the validity of the total procedure appears to have been higher than .45 No data are available to show the validity of any one test used in this program, although that would be essential to the evaluation and improvement of the procedures

A group of somewhat similar devices was tried out in the Aviation Psychology Program of the Army Air Forces (316 656–669,797 554–555), but in the regular testing of 400 cadets per day rather than in the intimacy of a week-end house party for a score of men The test situations were the Observational Stress Test developed by Glen Heathers, the Observations During Rest Period devised by the same psychologist, and the Interaction Test planned by the writer In the first the examinee was rated for promise as a pilot on the basis of his observed reactions when presented with a confusing multiplicity of stimuli while manipulating the controls of an airplane, in the second similar ratings were made while the cadet waited in a room furnished with a bomb, a twisted piece of fuselage, and other reminders of the dangers of military flying, in the third, the same type of rating was made (by a different psychologically trained enlisted man) while four cadets jointly assembled three Wiggly Blocks, a situation calculated to provide some opportunity to reveal leadership ingenuity, and ability to co-operate Only the ratings based on the Observational Stress Test and the Interaction Test had any validity when correlated with success in flying training, and these were so low as to be doubtful In addition to the overall ratings of pilot promise, specific ratings of co operation, leadership, emotional stability, and other traits were made during the Interaction Test, but these validities also were negligible (rbis ranged from −.17 to .13)

Perhaps the above-cited data only demonstrate that these specific forms of the situation test have no predictive value for this one type of behavior, success in flying training But other measures had validities ranging up to .51 (214 191) for this type of behavior It therefore seems clear that each such test should be specifically validated for the type of behavior it is intended to predict, until the remote day comes when they have been demonstrated to be such pure measures of traits, the respective

significance of which has in turn been proved valid for each specific type of vocational activity, that it is safe to generalize from test to occupation without a correlation coefficient to justify the prediction

Situation tests therefore appear to be promising techniques for the study of personality with its vocational implications, but their demonstrated validity is not at present such as to justify their use on any basis other than that of clinical intuition The underlying logic and their face validity suggest, however, that they should be experimented with in personnel selection programs and their validity established, particularly for positions of leadership and responsibility

Incomplete Sentences Tests In this technique the examinee is presented with a list of incomplete sentences or stimuli such as "I wish ," "My boss ," "The work I do ," and "My mother " The specific stimulus phrases vary with the purpose of the test, with the attitudes and traits which it is desired to assess They have been experimented with by Payne (595), Thorndike and Lorge (484), Rohde (642), Sanford (666), Tandler (817), Rotter and Willerman (655), and in an unpublished study of commercial airline pilots by Hobbs, the technique originates in the intriguing word-association technique developed by Jung, experimented with rather fruitlessly by many investigators (810) and most recently reviewed by White (921) The special advantage of this open-end or sentence-completion technique is the freedom which it leaves the examinee to reveal his true feelings by the way in which he structures a semistructured situation This complicates scoring, but devices are being experimented with for the categorization of responses in such a way as to make possible the rapid classification and scoring of the completed sentences Although there is little evidence as yet, what there is seems to suggest that the technique may develop into a method of measuring attitudes and needs which is more subtle and more valid than the attitude or personality inventory If so, it may prove useful in vocational counseling and personnel work when problems of job satisfaction and morale are likely to be important, and also in screening maladjusted persons for clinical counseling

CHAPTER XX

APPRAISING INDIVIDUAL VOCATIONAL PROMISE

Preliminary Considerations

Focus on the Individual

IN THE early chapters our attention was focused on the logic and steps of test construction and validation, on the nature and occupational significance of a variety of aptitudes and traits, and on instruments for their measurement As pointed out in the introduction, this focus was chosen because in actual work with tests one begins perforce with a test result, and proceeds to study the significance of that score for the occupational plans of the person being counseled Intimate knowledge of the construction and validation of each test used is essential to test selection and to test interpretation But in appraising individual vocational promise, whether in a counseling or in a personnel capacity, there are other steps which precede and follow the selection and interpretation of tests When the focus is on the individual rather than on a test the perspective changes and other considerations come to the fore For these reasons it is the purpose of this chapter to consider the use of tests in appraising individuals

What is said here does not bear on the work of the psychologist or personnel man who is using tests mechanically in large-scale selection programs, in such work the procedures are those of test development, described in another chapter, test interpretation is then simply the statement of chances of success as expressed in a numerical score For example, it was ascertained through test validation that an aviation cadet with a pilot stanine of 9 had 84 chances in 100 of being successful in flying training, whereas a cadet with a stanine of 1 had only 19 chances in 100 of succeeding in flying training (214 145) In such operations there is neither a problem of selecting appropriate tests nor one of synthesizing the results and evaluating their significance for a given individual, for test selection has been taken care of in the test development program,

and synthesizing results and teasing out meaning has been taken care of by the validation and scoring processes For a more detailed discussion of statistical interpretation, see Appendix A

But the material of this chapter is of importance to the worker who must operate without extensive previously validated test batteries It is important to most persons working in small organizations, with small departments, with executives even in large organizations, and as private consultants Many of the applicants appraised in these situations are considered for positions which have not been thoroughly studied with tests, and which sometimes cannot be so studied in time to help with the solution of immediate problems In such instances the user of tests must operate more as a counselor or clinican, bringing together bits of information about tests and about jobs in order to make the best possible appraisal

The material in this chapter is even more important to the vocational counselor whose function it is to help his clients to obtain the most accurate possible picture of their abilities and interests in relation to occupational opportunities In such work the counselor usually has to help the client do what he should have been doing for some years previous review his school, leisure-time, and work experiences in order to understand what they reveal concerning his vocational abilities As pointed out in an earlier chapter, vocational appraisal in counseling often requires the analysis of a much greater number of abilities, and the consideration of the requirements of a much greater variety of occupations, than does appraisal in selection work Needed occupational norms are often not available, and those that might be used are often for populations of such specialized characteristics as to make generalization to other seemingly related occupations a questionable procedure The use of tests by a vocational counselor is therefore of necessity generally not a predictive process but rather a clinical procedure A variety of data have to be studied in relation to each other, and hypotheses are established for the consideration of the client It should be noted that the term hypotheses is used, rather than conclusions, as their bases are not definitive enough to warrant the term conclusion The client decides which hypothesis seems most likely to him, aided by the mature experience and accepting attitude of the counselor, and proceeds to test it by embarking upon an appropriate plan This plan is subject to review and revision on the basis of subsequent experience, either with the continuing aid of the counselor or by the client alone

Selecting Appropriate Tests

When utilizing psychological tests for the appraisal of vocational promise, the first problem with which one is confronted is that of the selection of tests suitable to the person and purpose at hand Until all people have a uniform cultural and educational background, and a standard battery has been developed and validated for a great many occupations (should that time ever arrive!), this is no mean problem At least four considerations must be kept in mind in making the selection

The person to be tested must be understood The psychometrist or counselor selecting the tests must know certain obviously important facts such as age, amount of previous education and approximate intellectual level All of these affect, for example, the choice of the Kuder or the Strong interest inventories, age and intelligence, the choice of the O'Rourke or Bennett mechanical aptitude tests As has been well demonstrated by the investigations of social psychologists interested in race differences and by the experience of vocational counselors working with refugee groups, the cultural background of the client is equally important Even when there are no language differences, differences in experiences peculiar to a sub-culture can affect the appropriateness of a test It has been found, for example, that in a picture-completion test standardized on American children and depicting, among other things, a boy about to kick something which he has just dropped from his extended hands, Scottish children often make the mistake of giving the boy a pumpkin to kick instead of the oval football The reason is clear their football game is soccer, in which a round ball is used, and they are familiar neither with oval balls (despite English Rugby-football) nor with pumpkins Some of our other tests, designed by and standardized upon residents of the Northeastern and Middle Western states, are not fully applicable to those who reside in other parts of the country

The purpose of testing must also be clear to the selector of tests Is the objective a survey of the abilities and interests of the client, in order to ascertain which areas might profitably be explored either by tests or by life experiences? If so, a combination of tests which tap a number of fundamental abilities and interests is desirable, even though occupational norms may be defective, for the important thing is to locate strengths for further study Or is the aim to make an intensive analysis of some one or two areas, in order better to understand and evaluate the possibilities of assets already known to exist? In this case, a number of

tests measuring varied aspects or manifestations of the same aptitudes may be desirable, to make possible a detailed study of an area For example, a number of tests of manual dexterity may be used to determine just what type of hand-and-finger operations the client performs with the most skill, or several interest inventories may be administered so that discrepancies between patterns on tests constructed in different ways and using different types of items may suggest special outlets to be avoided or sought

Whether the testing is to aid in guiding development over an extended period or to help in making an immediate decision is another aspect of the purpose of testing A young man who has left school with no intention of continuing his education but who wants to get started in a field in which he may be able to learn and progress on the job is in quite a different position from another who expects to go to college and wants help in deciding what to major in and at what field to aim Directional guidance is sufficient for the latter case, and this calls for a variety of tests and inventories in order to check the level at which he may work and to point out occupations which he may do well to explore in courses, extra-curriculum, and summer jobs But, reluctant though one may be to work in such a way, the case of the young man deciding on an entry occupation requires careful study of qualifications for immediate employment The study must cover previous experience, and that is often very helpful, but in other instances test results are the most tangible and clear-cut guide available The battery of tests must therefore be one which throws direct light on qualifications for entering at once into any one of several occupations under consideration To fail to get all possible meaning from tests in such a case is to leave the making of the decision largely to chance

The vocational aspirations of the client are a third factor determining the selection of tests The psychometrist or counselor must know not only the background of the client and the nature of the service being rendered, but also the ambitions or goals which the client has in mind He must know what educational and occupational level he hopes to attain, as some aptitudes are more important at some levels than at others (e g , clerical perception), he must also consider what type of occupation the client hopes to enter, as that will help him decide how fully to test in special areas such as the technical and linguistic

Test data constitute the last type of information necessary in the selection of tests for use in counseling Knowing the client's status and goals, one must choose tests which have appropriate contents and norms, which

are known to measure traits relevant to the choices in question, which measure these reliably, and which can be administered and scored in the time available There should be no need to elaborate on these points in such a treatise as this

Three Methods of Vocational Diagnosis

Historically and currently, there are three methods of appraising the vocational promise of an individual with the aid of tests one is clinical and two are psychometric Their fundamental differences lie in the way in which tests are used In the *clinical* method the results of each test are viewed singly and in relation to other tests and to personal and social data All of these are weighted mentally, and a subjective judgment is made on the basis of this weighting In the *psychometric profile* method test scores and other quantifiable data are compared with occupational norms, as when an individual's test profile is plotted and visually matched with those of various occupational groups to ascertain which he resembles most clearly In the *psychometric index* method quantification is carried one step further to permit the expression of the individual's summarized test scores in one total score or index This shows how he compares with members of the occupation in question Thus in the Aviation Psychology Program of the Army Air Forces the scores of each cadet were statistically weighted and combined to yield three scores or stanines which expressed his standing as a prospective pilot (the pilot stanine) navigator (the navigator stanine), and bombardier (the bombardier stanine) These procedures are discussed at some length in the following sections

The Clinical Evaluation of Test Data

The clinical method of evaluating test scores was the first to be used in vocational counseling, because occupational data were not available to make possible the psychometric methods It has not often been described in the literature, perhaps because its very subjectivity makes it difficult to describe, a good recent discussion of test interpretation has been prepared by Harmon (332), emphasizing the profile method but including the clinical Its advocates are many, and there are many who claim that it is not only the first method to have been used but also the ultimate method, to which all will turn when the defects and limitations of the psychometric methods are more clearly understood This argument is met with the reply that, as psychometric methods improve, more factors will be more adequately taken into consideration and judgments made

subjectively by the counselor will be made objectively by psychometrics The reasoning underlying this statement is that anything that exists can be measured, and that any relationships which exist can be quantitatively expressed if the clinician can do it, science can do it more accurately The writer is inclined to agree with this latter position, but to recognize also that science must make a good deal of progress before all significant factors and relationships can be quantitatively measured and expressed For this reason the clinical method of test interpretation is of great practical importance and should be adequately described

The objective of the clinical method is to describe the individual in dynamic terms, in the expectation that a good picture of the person will make possible inferences concerning occupational success and satisfaction The underlying hypothesis is that genuine understanding of a person, combined with insight into a situation, permits one to foresee the interaction of forces and predict the outcome More humbly and accurately put, they permit one to set up hypotheses concerning the probable outcomes Even when stated in these terms, it is clear that the clinical method takes on no mean job and makes claims as great as those of the psychometric, perhaps even greater, for the best psychometric predictions are made with full consciousness of the limited basis upon which they are founded, whereas the clinical method attempts to take into account all that is revelant It puts great weight on the training, insight, and objectivity of the counselor

In the writer's experience as a counselor, supervisor of counselors, and counselor-trainer, there have seemed to be three principal techniques of ultilizing the clinical method These are the case conference, discussion with the client, and the preparation of psychometric reports Separate chapters are devoted to the last two topics, so they will be only briefly mentioned here

In the case conference the test scores are presented so that all in attendance may see them, generally on a blackboard Sometimes they are simply listed, and sometimes they are plotted in graphic or profile form The counselor orally summarizes the background information, giving the staff an outline of the socio-economic status, education, previous experience, interests, aspirations, and presented problem of the client The counselor or psychometrist then reviews the test scores, commenting on any observations made during testing that may add to the data The case is then thrown open for factual questions, after which members of the case conference raise questions of interpretation, propose interpretations of their

own, and make suggestions for further investigation or counseling At the close of the conference the counselor or chairman summarizes the discussion, perhaps attempting to present an integrated picture of the case as seen by the conference The focus may be on diagnosis, but in practice it generally includes also the nature of the counseling and the resources which may be utilized in implementing the counseling

Case conferences such as these are unfortunately rarely held in service agencies other than hospitals and special institutions, largely because of the amount of time they require They are common in training situations, whether academic or institutional, and some service agencies make a practice of holding them occasionally as an in-service training or supervisory device They have a number of advantages as a clinical diagnostic technique 1) they utilize the insights and resources of more than one counselor, 2) they are a safeguard against blindspots and biases, 3) they force the crystallization of ideas which might otherwise not be made clear and concrete

Discussion with the client resembles the case conference as a technique, but with the important difference that at least one of the discussants is untrained in the use of tests and is emotionally involved in the proceedings Despite these facts such discussion does a great deal to clarify the counselor's thinking about the significance of the test scores, partly because of the freshness of another person's point of view, and partly because the opportunity to think out loud brings ideas to the surface Furthermore, the client's reactions to the data and to the counselor's tentative interpretations (often put in the form of a question beginning with "Could that mean ?") provide a healthy corrective for the counselor's own possible biases This procedure is discussed at great length in the next chapter, from a somewhat different viewpoint

The preparation of test reports is perhaps the commonest and best technique for the application of the clinical method of test interpretation In writing up the results of testing the counselor not only expresses the test scores in verbal form, but discusses the significance of background data, observed behavior, and client attitudes and statements for the interpretation of test scores, relates test scores to each other and to these non-test data, and draws conclusions concerning the true characteristics of the individual being studied These are then related to each other in a final summary or thumb nail sketch of the client as seen through the interpreted test data This process, like the others described, forces the counselor to crystallize his ideas and to justify his interpretations, at least in

his own eyes and in the eyes of any potential reader It thus ensures more thorough exploration of the data than would a mere mental interpretation of test scores, and provides something of a safeguard against the indulgence of bias and the riding of hobbies This technique also is treated at greater length and for a different purpose in a later chapter

The picture of a person obtained by the above methods is probably as adequate as any The interpretation of test data and case-history material, if the data themselves are skillfully obtained, is in fact the only method available for the psychological description of an individual But from the point of view of vocational counseling, the defects of the clinical method are two 1) the evaluations or judgments made are subjective, so that even a group of experienced counselors may be wrong, and 2) the best techniques for describing the psychological characteristics of an individual may be lacking in data concerning their occupational significance

In the occupational applications the judgment of the counselor again becomes of fundamental importance One might cite the O'Connor Tweezer Dexterity Test as an example, interpreted for years as a measure of significance for success in dental school, but shown by the majority of studies to have doubtful validity for that purpose Or reference might be made to Thurstone's work with primary mental abilities tests, which there is every reason to believe measure basic human aptitudes but which have not yet been actually demonstrated to have occupational significance Several wartime aviation psychologists were certain that they could make better predictions of success in flying training by clinical interpretations of test data than were provided by the objectively obtained stanines, but, either because of the inadequacy of some of the tests used or because of their lack of knowledge of flying, or both, their predictions had no validity (316 669,616,797) As it is known that many instruments are good measures of psychological characteristics of one kind or another, and relatively few have been validated for many occupations, it is probably in the making of occupational applications that the clinical method makes the gravest errors It is one which should be used only by counselors who have acquired both an intimate knowledge of tests and an even greater fund of information concerning occupational activities and requirements

Methods of drawing on a general fund of occupational information for the clinical interpretation of vocational tests, and of adding to that fund when it is not sufficiently great or detailed, deserve some mention they

are even less frequently treated in the literature than are methods of analyzing test results in order to prepare a psychological sketch of an individual They amount to the making of a job analysis by the psychometrist or counselor If he has a good fund of vocational information the job analysis is of the armchair variety the counselor mentally reviews the functions, duties, and tasks of workers in the occupation, and makes deductions concerning the aptitudes and traits which seem to be required, he checks these deductions against what he knows of the published material on test validities and against the expressed opinions of others who are familiar with the work in question The list of characteristics thus drawn up in his mind, and perhaps put on paper, serves as a guide in considering the client's qualifications for work of the type in question

If the counselor lacks sufficiently detailed information concerning the occupation in question, the job analysis must be made from a vantage point other than the armchair The first step may be familiarization with printed material in the form of occupational and industrial descriptions such as are listed in Shartle (714) and in Forrester (263) But such data are often too general to provide the insights needed into the aptitudes and traits which make for success on the job The counselor then needs to go to the job itself, observing workers in action, familiarizing himself with the knowledge, tools, processes, and problems of the occupation This takes time, but it is the accumulation of information acquired in such first-hand contacts with vocations and workers which distinguishes the vocational counselor from the clinical psychologist The latter knows diagnostic and counseling techniques, and has insight into the dynamics of human adjustment, but unless he has had a great deal of contact with workers and has studied their work he is not qualified to do vocational counseling The techniques used in these field studies and observations are of course the standard techniques of job analysis as used in the preliminary work of test development Shartle (714) has described them in his text on the collection and organization of occupational information

The Psychometric Profile Method

The first attempts to objectify the clinical method of test interpretation consisted of administering batteries of tests to persons in a variety of occupations in order to ascertain the nature of the patterning of test scores This was the method developed by the Minnesota Employment Stabilization Research Institute (223,589), which used a standard battery of intelligence, clerical, mechanical, spatial, and manual dexterity tests,

administering this battery to groups of clerical workers, department store clerks, policemen, janitors, accountants, casual laborers, and others The mean scores made by each group on each test were ascertained, and a profile plotted for each group, as shown in Figure 8 This made it possible to give the same battery of tests to a client, and to compare the patterning of his scores with that of accountants if he aspired to be an accountant, or to the patterning of the aptitudes of policemen if that was an occupation to be considered

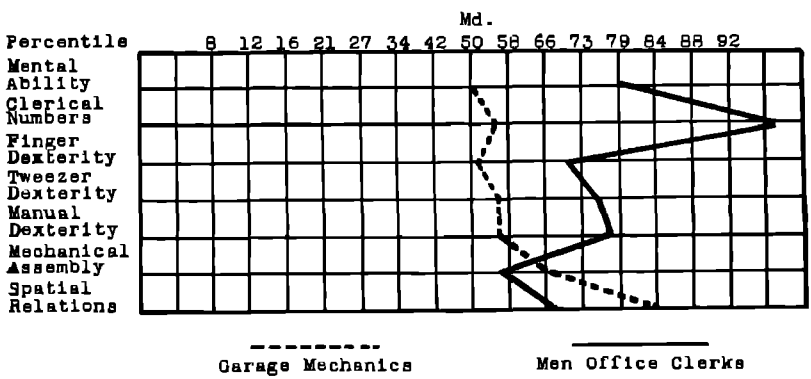


FIGURE 8

OCCUPATIONAL ABILITY PATTERNS
After Andrew and Paterson (22)

The technique had a number of serious limitations, of which its originators were well aware One was the limited number of occupations for which patterns could be obtained, this was in part remedied by the generalizations of experts in the *Minnesota Occupational Rating Scales* (591) Another was the difficulty of deciding when an individual's profile differed significantly from that of an occupational group, discussed in connection with the USES General Aptitude Test Battery, this was then remedied only by the judgment of the counselor, making the method partly clinical in nature A third was the limited number of characteristics appraised by the test battery and included in the profile, this also had to be remedied by the counselor's clinical skill and occupational knowledge As in the case of the clinical method, too often counselors have knowledge of tests or knowledge of occupations without having both Finally, the populations used to establish occupational ability profiles in the Minnesota project were selected as representing the local population, leaving the question of their applicability in other localities unanswered

The Occupational Analysis Division of the United States Employment Service carried work with this technique further, partly for selection and partly for guidance purposes In the former program batteries of the most valid tests were used in varying combinations to establish profiles for each job studied, in the latter, a standard battery was administered to persons employed in various families of occupations and patterns of aptitudes were ascertained This work has been described in Chapter 15, its outcome being the USES General Aptitude Test Battery The difficulties discovered in the MESRI work were minimized in the USES project by classifying occupations in families in such a manner as to make some 200 profiles represent approximately 2000 major occupations, basing the profiles on critical minimum scores rather than mean scores, selecting the tests for inclusion in the battery on the basis of a factor analysis of vocational aptitudes, and sampling occupations in various key parts of the country rather than in one or two localities As was pointed out in the discussion of the tests, the battery still has defects, but it represents a great advance in the occupational ability pattern or psychometric profile method Its usefulness is limited, however, to the tests used in the original battery (not available except to the state employment services) and to the occupations already studied It makes one further contribution, in that a counselor who knows the patterns established by the General Aptitude Test Battery, and who has a real understanding of vocational processes and requirements, can use this fund of information to provide an objective foundation for the exercise of clinical insight when working with tests and occupations for which occupational ability pattern data are lacking

The Differential Aptitude Tests of the Psychological Corporation, also described in Chapter 15, are another attempt to improve and extend the occupational ability pattern or psychometric profile method, although to date it has been applied only to school populations *The American Institute for Research* has in preparation a comparable battery, based on the wartime studies of aviation psychologists, and other such batteries are also being planned (320)

The usefulness of this method of appraising vocational promise depends, as might be expected in the case of an empirical method, upon the accumulation of objective evidence It has been seen that only a bare minimum of such data are now on hand, enough to reveal the promise and the defects of the technique, to provide a concrete basis for the making of some decisions, and to make somewhat less intuitive some of the

clinical judgments which have to be made when objective data are lacking It would be pertinent to ask whether there may not be real danger of a too mechanical application of this psychometric method once more occupational ability patterns are available, for it is certain that no test battery in the foreseeable future will be able to measure every trait which has a bearing on success and satisfaction, especially when it is remembered that some of the factors which determine success and failure are not personal or psychological, but rather environmental or economic and social But discussion of this question is postponed until the end of the next section

The Psychometric Index Method

The combining of test scores in order to provide a single score or index of vocational promise has long been practiced, both in the arbitrary weighting of scores on the basis of *a priori* judgments of their relative importance in a job, and in the statistical weighting of test scores on the basis of their respective correlations with the criterion This has been a selection technique, however, rather than a method of appraising an individual for counseling, largely because data were lacking for the statistical weighting of tests for counseling and because counselors were properly reluctant to give the appearance of objectivity to their judgments by arbitrarily weighting the scores and combining them Perhaps two exceptions to these statements may now be made

In the Kuder Preference Record (pp 445 ff) a series of scores (those for the nine types of interests) can be weighted on the basis of their relationship to membership in an occupation, and combined to show how closely an individual's interests resemble those of members of that occupation This is a limited application of the technique, both because the scores involved represent only interests, and because such occupational indices, as Kuder calls them, have so far been developed for only two occupations

The stanines of the Air Force's Aviation Psychology Program (214) may be a second exception Although these were developed for selection purposes rather than for counseling, the fact that they are available for three different flying jobs and are all obtained from the same basic test battery means that they could also be used in counseling concerning the choice of any one of those three specialities This too is a limited application of the technique, but it illustrates its possibilities

The fundamental argument for the use of the psychometric index is that it does away with the subjectivity of the profile, instead of leaving

to the counselor the making of an overall judgment of the similarity of the client's psychological characteristics to those of members of the occupation in question, this "judgment" is made by an empirically based statistical process which is more precise than the subjective judgment of the counselor Each aptitude and trait is weighted on the basis of its occupational significance, and the similarity of the individual to others who have succeeded in that occupation is expressed by the final score or index In the Air Force, for example, a cadet with a stanine of 9 is known to have aptitudes, interests, and temperament very much like those of other cadets, 84 percent of whom succeeded in learning how to fly, while another cadet with a stanine of 1 is clearly shown to have characteristics like those of cadets, 81 percent of whom failed to learn to fly in the alloted time (214 145)

As was mentioned in the discussion of the clinical method, this procedure has sometimes been criticized as too mechanical, as failing to take into account the multiplicity of personal and social factors which affect success and satisfaction Probably no one would contend that it does take all of these into account its proponents would argue only that what it does consider is taken into account in the most accurate manner possible The adequacy of that kind of appraisal is a matter, not for discussion (except for the establishment of hypotheses), but for experimentation Two kinds of evidence are available, one a comparison of the effectiveness of various clinically appraised test data with that of mechanically computed Air Force stanines for the prediction of success in training, the other a comparison of the effectiveness of clinically evaluated and mechanically applied stanines for the same purpose Both of these are experiments in a selection rather than a counseling situation, unfortunately the lack of psychometric indices for use in counseling has precluded the possibility of making such experiments in counseling programs

In the *Clinical Techniques Project* of the Army Air Forces, aviation psychologists experimented with a number of clinical evaluation tests for the selection and classification of pilots (316 Ch 24,616) These tests included ratings of prospects of success based on observations made while the cadet responded to a confusing sequence and combination of signals in a miniature cockpit, while he worked with three others to assemble the parts of three sets of Wiggly Blocks, and while he sat in a waiting room surrounded by odds and ends of wrecked airplanes, bombs, and similar objects, they also included the Group Rorschach, which was scored in the usual manner and also evaluated impressionistically to yield a rating

of promise as a flier As has already been mentioned, none of the techniques had any substantial validity, and those which showed some slight promise in the first validation were proved invalid in the cross-validation At the same time, the objectively derived stanines had their usual substantial correlations with success in pilot training It should be pointed out that this was a very limited evaluation of the clinical method, for each clinical evaluation was based on only one source of data, however global in approach the test was. Although it had been planned to make evaluations on the basis of a clinical synthesis of all data for each cadet, this part of the plan broke down because of the sheer bulk of the data to be handled and the impossibility of assigning the required number of psychologists to the project over such a long period of time

The Surgeon's Classification Board provided an opportunity for a more comprehensive clinical evaluation of cadets being considered for flying training during several months in which it was experimented with during World War II (described in a military report by W M Lepley and H D Hadley) The board consisted of a flight surgeon and an aviation psychologist, who interviewed each cadet with stanines below the required levels for all three air crew assignments (at that time 3 for pilot and bombardier, 5 for navigator) The interviews lasted approximately eight minutes each, ranging in length from five to twenty minutes A total of 1524 cadets were interviewed during the six months of the board s existence at this one classification center, and 285 were sent to pilot training because the board s review of the test scores and interview data led it to believe that the cadet would make a good pilot Follow up data were obtained for 259 of these cadets, who were test-matched with 146 cadets sent to training at a somewhat earlier date when standards were lower and without having been passed on by a board Various analyses were made by class and time of training, in the most legitimate comparison, 68 9 percent of the cases passed by the board failed in training, whereas 73 percent of those with similar stanines who went automatically to training failed The critical ratio was 0 50, showing that cadets who were clinically evaluated by a board of experts were no more likely to succeed than others who had the same stanine or psychometric index but were not clinically evaluated Despite certain defects in the design of this real-life experiment (e g , elimination rates were not quite the same when the two groups were in training, being slightly lower for and therefore favoring the board cases) Lepley and Hadley seem to have definitely put the burden of proof upon those who claim that the clinical method is

superior to a comprehensive battery of objectively validated and summated tests At present, one can only conclude that the rather superficial but costly clinical methods which have been evaluated have been proved no more effective than the less time consuming objective methods

A Balanced Approach in Counseling

The preceding sections have brought out the facts that use of the clinical method is often necessitated by the lack of data basic to the use of psychometric methods, and that the fully developed psychometric index method is not easily improved by adding clinical evaluation to it It has also been made clear that both methods depend for their success on the use of a variety of relevant and well-understood tests In view of the scarcity of psychometric occupational indices, the clinical method, made as objective as possible by occupational norms, must generally suffice as the technique of individual appraisal for vocational counseling

In closing this discussion, a word of caution needs to be put on record concerning the mechanical use of test results The presumed superiority of completely validated and objectively summated test data over clinical interpretations does not mean that test results should be used mechanically, if "mechanically" is taken to mean applied indiscriminately and regardless of the background of the person taking the tests, his health and morale at the time of testing, and the conditions of testing Clinical interpretation in this sense is always necessary, and in counseling it should be easier than in a large-scale selection program One illustration will perhaps suffice to make the point It will be remembered from the discussion of test administration that Meltzer (524) reported a correlation between manual dexterity and output in an industrial job which changed from $-.27$ to .30 with a change in supervision The attitudes of the persons taking the tests and producing the output are important When such factors are involved the clinical insight of the test user is crucial He cannot know whether or not such factors are present unless he has insight and is alert to use it

CHAPTER XXI

USING TEST RESULTS IN COUNSELING

THE interpretation of the results of psychological tests, whether by the counselor for his own diagnostic purposes as discussed in the preceding chapter, or for the counseling of clients as considered in this chapter, has been strangely neglected by most authors of books or articles on the use of tests in guidance In the texts of the mid-thirties there was some mention of problems of technique, but it is only with the focusing of attention on interview techniques which resulted from the work of the nondirective school that it has been written about in detail In view of the relative recency of some of these developments and the controversy which still surrounds them, it seems wise to describe the techniques of transmitting test results to clients as they have been reported in the literature before attempting to suggest a method which combines the strengths of several

In a treatise such as this it is difficult to observe the distinction between test interpretation and counseling, indeed, it could be maintained that there is none, for test interpretation is one technique of counseling But it is only one technique, and a very limited one, despite the fact that some psychologists who are more skilled in psychometrics than in counseling have acted as though it were the principal method of counseling As a technique, it can legitimately be singled out for discussion by itself, it must be remembered, however, that it can be fully understood only within the framework of counseling in general Chapter I has been devoted to a discussion of counseling, in this chapter, the focus is therefore as narrow as possible on test interpretation With this caution in mind, we may proceed to survey methods of test interpretation

Directive Test Interpretation

One of the first specific discussions of interpreting test results to clients appeared in 1937 in Williamson and Darley's *Student Personnel Work*

(931) After describing the types of material included in the synthesis of test and other personal data, they wrote "It is the job of the counselor to integrate this material, to interpret the present abilities and achievements of the case in terms of his background, and to draw conclusions from these interpretations The final act of counseling the case is not performed by instructing the student to train for this or that particular profession, but by presenting to the student his possibilities in certain lines of endeavor, alternative goals, with the evidence for and against a choice, help to clarify the student's thinking and provide needed data for a tentative decision He is urged to try, at least, that course which seems to suit his abilities and interests most favorably, the tentative nature of the try-out and the necessity for further interviews, before a final decision, are emphasized" (931 166) Again "The *recommendations upon which prognoses are based must be in terms of alternatives so that the student may make his own choice* It is at this point in the case work that the counselor translates his two basic principles, about prediction for success in training and prediction based upon the characteristics of goal groups or occupational groups, into terms that the student can understand in relation to his own problems" (italics in the original) (931 175) [1] Williamson and Darley gave no more space to test interpretation in this book, but this brief discussion makes it clear that they viewed the process as one explaining logically and in everyday language the significance of tests and their vocational implications to the client

These points were elaborated upon somewhat in later books by the same authors In *How to Counsel Students* (928) Williamson wrote "The counselor must begin his advising at the point of the student's understanding, i e , he must begin marshaling, orally, the evidence for and against the student's claimed educational or vocational choice and social or emotional habits, practices, and attitudes The counselor uses the student's own point of view, attitudes, and goals as a point of reference or departure He then lists those phases of the diagnosis which are favorable to that point of reference and those which are unfavorable Then he balances them, or sums up the evidence for and against, and explains why he advises the student to shift goals, to change social habits, or to retain the present ones The counselor always tells what a relevant set of facts means, i e , their implications for the student's adjustment, in other words he always explains why he advises the student to do this or

[1] By permission from *Student Personnel Work*, by Williamson, E G , and Darley, J G Copyrighted 1937 McGraw Hill Book Co

that, and he does the explaining as he orally summarizes the evidence If in this way the student's confidence in the counselor s integrity, friendliness, and competence has been secured, the student should be ready to discuss the evidence and to work out cooperatively a plan of action" (928 135) [2] Although there is little mention of tests in the preceding material, it is clear that in a University Testing Bureau such as Williamson directed much of the evidence presented to the student would be in the form of test data After a survey of other methods of counseling, and with only a passing reference to passive or indirect methods (as the nondirective were then called), Williamson took up the *explanatory method* in more detail "In using this method the counselor gives more time to explaining the significance of diagnostic data and to pointing out possible situations in which the student's potentialities will prove useful *This is by all odds the most complete and satisfactory method of counseling* [italics original], but it may require many interviews With regard to vocational problems the counselor explains the implications of the diagnosis (of test and personal data) and the probable outcome of each choice considered by the student He phrases his explanation in this manner

" 'As far as I can tell from this evidence of aptitude, your chances of getting into medical school are poor, but your possibilities in business seem to be much more promising These are the reasons for my conclusions You have done consistently failing work in zoology and chemistry You do not have the pattern of interests characteristic of successful doctors which probably indicates that you would not find the practice of medicine congenial On the other hand you do have an excellent grasp of mathematics, good general ability, and the interests of an accountant These facts seem to me to argue for your selection of accountancy as an occupation Suppose you think about these facts and my suggestion, talk with . , see , and return next Tuesday at 10 o'clock to tell me what conclusion you have reached I shall not attempt to influence you because I want you to choose an occupation congenial to you But I do urge that you weigh the evidence pro and con for your choice and for the one I suggest' " (928 139–140)

In *Testing and Counseling in the High School Guidance Program* (190 Ch 7) Darley wrote in the same vein The counseling interview, he stated, may be thought of as an unrehearsed play in which the coun-

[2] By permission from *How to Counsel Students*, by Williamson, E G Copyrighted 1939, McGraw Hill Book Co

selor "carries" the action, a special learning situation for the student, a cathartic experience for a student suffering from great emotional pressure, or a sales situation In all but the cathartic type of interview Darley conceived of the counselor as taking the lead, for in the play he must "organize the conversation" and "summarize the action", in the learning situation he "explains the assembled test material and non-test data to the student, and then follows this by a discussion of the material" (190 169), and in the sales situation he "attempts to sell the student certain ideas about himself, certain plans of action, or certain desirable changes in attitudes Persuasion and logic will facilitate and hasten the sale of such ideas by a counselor" (190 169) Darley continued (190 179) "Many books on guidance insist that the counselor must not *tell* the student what to do While such a generalization seems unsound, since it emasculates most of the purpose of data collecting and since it would be of no assistance to a student who needs help in making a decision, it is still true that the student who chooses one from among several suggested plans of action will feel a more active participation in planning with the counselor "[3]

Experience and the contributions of others may have led both Williamson and Darley to modify their viewpoints since the above texts were written, for much progress has recently been made in the clarification of counseling methods, but these writings have influenced and are influencing many users of tests and many counselors For this reason it is necessary to present them in some detail Perhaps the best way to see the limitations of this method is to present the antithetical point of view, and then to attempt a synthesis

Nondirective Test Interpretation

Most active and severest critics of the directive approach of Williamson, Darley, and many other vocational psychologists and counselors have been the nondirective counselors, led by Rogers (640,641) His most detailed discussion of the use of tests in nondirective counseling (640) first points out that tests do not stand up well as a "client-centered" counseling technique, because in suggesting tests the counselor implies that he knows what to do about the client's problem, in administering them routinely or early in a contact he proclaims that he can find out all about the client and tell him what to do, and in interpreting them he

[3] By permission from *Testing and Counseling in the High School Guidance Program*, by Darley, J G Copyrighted 1943, Science Research Associates

poses as an expert who knows all the answers and will impart them to the client "By every criterion, then, psychometric tests which are initiated by the counselor are a hindrance to a counseling process whose purpose is *to release growth forces* They tend to increase defensiveness on the part of the client, to lessen acceptance of self, to decrease his sense of responsibility, to create an attitude of dependence upon the expert" (italics are mine) (640 141) As one might expect from the italicized clause, however, Rogers went on to point out that there are stages in counseling at which clients are emotionally ready to study their abilities and interests and to compare them with those of others as a part of the formulation of objectives and the making of plans Rogers believes that this does not occur frequently in practice, and that it is not the factual test results which are important, but rather the attitudes of the client toward them He therefore sees little place for tests in practice while admitting that there is one in theory

Rogers' views may have been conditioned unduly by selective experience His theories were formulated while working in a child guidance clinic After that his work in a university counseling center undoubtedly brought him cases in which emotional problems were much more common and more serious than they are in cases going to the vocational counseling center of the same university It is a commonplace that people are referred to and gravitate toward persons who are interested in their types of problems psychoanalysts encounter sex problems, ministers religious problems, attitude (nondirective) counselors attitudinal problems It is significant that those of Rogers' students who have worked in centers which specialized more in vocational and educational counseling have found their viewpoints modified by that experience It is from them that some very helpful formulations of the role of tests in counseling and of the methods of interpreting test results to clients have come Contributions from the Bixlers, Combs, and Covner are cited below

The use of nondirective *techniques* in vocational counseling was considered by Covner (173) in a review of his experience in a vocational counseling center Concerning tests he wrote "A second major locus of fruitful application of the nondirective approach was the area of *preparation for testing and interpretation of test results* Test interpretation called for all the skill the counselor could muster As an introduction to interpretation it was frequently found helpful to sound out a client on his reactions to the tests His mode of response served as a guide and warning to the counselor as to what sort of session test interpretation

would be For example, when a client who did very poorly on certain tests reported that he 'knocked them for a loop,' the counselor took notice to proceed with caution The same approach on a number of occasions showed that clients were able to do a remarkably accurate job of interpreting their relative strengths and weaknesses, and to reveal considerable understanding of themselves" (173 71–72) Covner goes on to point out that rejection of the counselor's interpretations often seemed to be the result of failure to give the client sufficient time to react, and that exploration of client reactions was facilitated by reflecting feelings, as in the statement, "The results are rather disappointing to you" To the experienced vocational counselor who has not been unduly influenced by highly directive writings such as those of Williamson and Darley, these insights into test interpretation do not seem very surprising such nondirective *techniques* have been the stock-in-trade of good vocational counselors since the origin of modern vocational guidance, but because of greater interest in occupational information and in the counselor's use of tests, they simply have not been written up

Types of problems best handled by directive and by nondirective techniques have been analyzed by Combs (166) who worked in a university counseling center in which educational and vocational problems outnumbered emotional adjustment cases by three to one (322) One type of case best handled nondirectively was that in which the level of aspiration is definitely higher than demonstrated ability or in which there is a wide discrepancy between expressed and measured interests Combs points out that such discrepancies are warning signals, and that the emotion likely to be aroused by being brought face to face with them is best handled nondirectively He does not elaborate on method in this connection, but it consists primarily of accepting the client's feelings and of reflecting them in such a way as to make it possible for him to discharge the emotion, to accept himself, and then to discuss the situation and its implications objectively This subject has been most adequately covered by the Bixlers, whose contribution is discussed below

The Bixlers served as counselors in the Student Counseling Bureau (formerly the University Testing Service) of the University of Minnesota, in which the test-oriented philosophy of Williamson and Darley tended to prevail, they therefore made a point of studying the use of nondirective techniques in vocational test interpretation (97,98) which, strangely enough in nondirective counselors, they seem to consider synonymous with vocational counseling They begin by pointing out that there are

two aspects to the problems of test interpretation 1) presenting test results to the client in such a way that they are understood, and, 2) dealing with the client in such a way as to facilitate his use of the information Although they do not so state, it seems to the writer that Williamson and Darley focused on the former and did it rather well, except that, as implied in Covner's report, they may not have allowed the client to react to the presented facts enough to guarantee an understanding of them They appear to have failed to deal with the client in such a way as to ensure his being able to use the facts, depending entirely on his being sufficiently well adjusted emotionally to assimilate a mass of personal and therefore emotionally toned information As Rogers pointed out, this is sometimes the case, perhaps more often than Rogers recognized, but it is certainly not always so As the Bixlers put it "The grading of examinations at the end of the quarter verifies the ineffectiveness of books and lectures in giving information to students Vocational test interpretation is much more personalized and there is greater opportunity and reason for the student to distort or disregard information given to him" (98 147) How many innocent counselors have not been shocked when clients or former clients reported that "You told me my tests showed that I would make a good personnel manager " merely on the basis of one Strong's Blank score? The rules evolved by the Bixlers for interpreting vocational tests to clients are given below

1 *Give the client simple statistical predictions based upon test data* Examples "Eighty out of 100 students with scores like yours on this test succeed in agriculture" "You have more of this type of aptitude than 65 out of 100 successful accountants" This can of course be elaborated upon

2 *Allow the client to evaluate the prediction as it applies to himself* After merely stating the facts the counselor pauses, perhaps longer than he feels he should, in order to let the client react to the facts

3 *Remain neutral towards test data and the client's reactions* The counselor expresses no opinions, gives no advice, but in a warm and respectful manner listens to what the client has to say This is called acceptance, it is not the same as agreement

4 *Facilitate the client's self-evaluation and subsequent decisions* The counselor recognizes and reflects the feelings and attitudes of the client Example "You expected this, but it's hard to take" This makes it easier for the client to explore his feelings further, to release any related tensions, and to view the data and their implications more objectively

5 *Avoid persuasive methods* The counselor need provide no artificial motivation the test data and the exploration and release of related feelings should do that If they do not, neither will the exhortations or cajolements of the counselor

Some sample excerpts from cases are given by the Bixlers (98 151–152) one of these is reproduced and commented on below, in order further to illustrate the technique

C1 There are studies which demonstrate that students' ranks in high school, along with the way in which they compare with other entering students in mathematics, are the best indication of how well they will succeed in engineering Sixty out of one hundred students with scores like yours succeed in engineering About eighty out of one hundred succeed in the social sciences (names several) The difference is due to the fact that study shows the college aptitude test to be important in social sciences, along with high-school work instead of mathematics

S1 But I want to go into engineering I think I'd be happier there Isn't that important?

C2 You are disappointed with the way the test came out, but you wonder if your liking engineering better isn't pretty important?

S2 Yes but the tests say I would do better in sociology or something like that (Disgusted)

C3 That disappoints you because it's the sort of thing you don't like

S3 Yes I took an interest test, didn't I? (C nods) What about it?

C4 You wonder if it doesn't agree with the way you feel The test shows that most people with your interests enjoy engineering and are not likely to enjoy social sciences—

S4 (Interrupts) But the chances are against me in engineering aren't they?

C5 It seems pretty hopeless to be interested in engineering under these conditions, and yet you're not quite sure

S5 No, that's right I wonder if I might not do better in the thing I like—Maybe my chances are best in engineering anyway I've been told how tough college is, and I've been afraid of it The tests are encouraging There isn't much difference after all—being scared makes me overdo the difference (He decided to go into engineering and seemed at ease with his decision)

Several good features of this interview illustrate points made by the authors, and are worth pointing out, together with some defects in the use of the technique The first statement of the counselor (C1) is a factual statement of an actuarial sort without explicit personal applications of evaluations In these respects it is good It fails, however, to achieve

simplicity and clarity What kind of score or scores are referred to in the "sixty out of one hundred" sentence high school rank, mathematics grades, or mathematics achievement test scores? If the last is meant, is that the score which predicts success in eighty out of one hundred students in the social sciences? The last sentence in the paragraph suggests that a scholastic aptitude test is the predictor for social sciences, a mathematics test that for engineering This may have been made clear in unreported passages preceding this one, but as it stands the paragraph has to be carefully analyzed in order to be understood It need be no longer to be clear by itself

The client's first response (S1) illustrates the value of the method in obtaining free expressions of the client's feelings on the matter, and the counselor replies (C2) by recognizing and reflecting the feeling This causes the client to bring up the counter argument himself (S2), putting him rather than the counselor in the position of the weigher of evidence, the client is taking responsibility for working out a solution The counselor helps him continue the thought process by again reflecting feeling (C3)

The client pursues the matter further and it occurs to him that another test he took might throw some light on the matter (S3) This natural introduction of test results and discussion of them as one more bit of evidence is one of the very real strengths of this technique it should be noted that the client is putting the test data to use with the help of the counselor, rather than the counselor telling him what their implications are A follow-up of such cases would probably show much less distortion of test data and of counseling by the client, for the client has done his own thinking and reached his own conclusions in the presence and with the help of the counselor, rather than after he left the counselor's office

The counselor then reports the relevant test results (C4), reflecting feeling in the process in order to help the client clarify his thinking The client continues to keep control of the diagnostic process, interrupting (S4) to make a tentative interpretation for the counselor to check The counselor reflects feeling (C5), rather than repeating statistics The result is a summary weighing of evidence by the client, in which he reaches a conclusion based on an understanding and acceptance of his own limitations and an awareness of the assets which he may draw on to help carry out his plans

A Synthesis of Suggestions for Interpreting Test Results to Clients

The preceding sections have made it clear that the nondirectivists who have worked in vocational counseling have made a significant contribution to the literature on test interpretation Their insistence on analyzing counseling experiences and techniques has led them to formulate principles and to describe the use of methods of test interpretation which have long been in use by many counselors In verbalizing what is done they have crystallized thinking on the subject and thereby helped to improve practice As they approach the problem from the point of view of a systematic school of thought they have made some unique contributions in pointing out the implications of various interpretive techniques for counseling, they also run the risk, as shown in Rogers' paper on the subject, of failing to see other implications and other possibilities because of theoretically produced blindspots There are values also in some of the more directive procedures, and occasions when they are more effective For these reasons the writer prefers a more eclectic approach If the suggestions outlined in the paragraphs which follow appear more nondirective than directive, that is because the writer's philosophy and and approach, like those of Dewey, Kilpatrick, Kitson, Brewer, Taft, Allen, Roethlisberger, Cantor, and Rogers, are client-centered and nondirective, even though not Client Centered and Nondirective (that is, not those of a "school" of counseling) The writer's philosophy is also pragmatic, in that he is willing to use whatever works, and does not feel compelled to use only the techniques which are compatible with a system, valuable though systems are as means of making one conscious of the implications of the procedures used

Structuring the Counseling Relationship It may seem odd to begin a discussion of methods of test interpretation with a consideration of the structuring of the counseling relationship, but experience has repeatedly shown that the client's attitude toward test results is an important factor in his first contacts with the counselor, and that what happens in these first contacts makes it easy or difficult for the counselor to use the test results constructively in counseling As Bordin and Bixler (113), Covner (173), and others have put it, most new clients feel that their problems will be solved by the counselor and that tests will play a major part in the process, and many are confused as to what vocational counseling is like One prerequisite to good test interpretation, then, is the establishing of an appropriate mental set in the client, this is generally

referred to as structuring the relationship The techniques are partly verbal, partly nonverbal

Verbal structuring may be done by asking the client what kind of help he wants the counselor to give him and, if (as often happens) the reply is, "Give me some tests to tell me what I will succeed in," by replying, "You feel that tests will solve your problem for you " Most clients react negatively to this type of bald but accepting statement by another of what they have actually been thinking, it brings to the surface the realization that they must assume more of the responsibility for their actions, and that tests are not likely to provide any such clear-cut answers For the client to formulate and express these ideas himself is much more effective than for the counselor to do so for him the former constitutes the achievement of insight, while the latter may be no more than indoctrination Verbal structuring is also accomplished by an explanation of the counseling procedure, to make it clear that it consists of two persons, one of whom is trained in counseling and in occupational information, discussing the other person's aspirations, status, abilities, interests, and plans, surveying relevant facts, and considering their implications It brings out the fact that testing is one way of getting some types of data, that there are other types of data and other ways of obtaining them and that discussion is the crucial process

Nonverbal structuring is based on the old adage that actions speak louder than words If the counselor creates a permissive situation, acts as though he were interested in the client, and "accepts" the client's expressions of feeling, the client will generally sense that discussion is the essence of the vocational counseling, and that his own participation is the essence of the discussion, he will usually welcome the opportunity to make a genuine exploration of his vocational aspirations and status and will assume responsibility for working actively with the counselor in this enterprise Once this type of relationship is established, the counselor need have no fear that the use of tests will unintentionally result in the imposition of a vocational prescription on the client

Test Administration and Interpretation In the chapter on test administration attention was devoted exclusively to the problems and techniques of administering tests to individuals and to groups But what has been said about interpreting test results to clients has made it clear that it has broader implications, for the way in which testing is done has an important effect on the client's expectations of tests As Rogers (640) pointed out, the routine administration of tests or the giving of tests

early in the counseling process, i e, in an unstructured relationship, implies both that the counselor knows what to do about the problem and that he can find out what he and the client need to know by means of tests Such test administration is, in fact, a nonverbal or behavioral structuring of the situation The antidote is not necessarily, as Rogers implies, to refrain from testing early in the relationship, it may consist of so structuring the relationship by discussion and behavior as to make it possible to test without creating this undesirable mental set As clients often come with it already partly established, to do so requires special attention to the problem and a degree of skill, but it can be done The essential factor is that testing be done by mutual agreement for jointly established purposes Ways of accomplishing this are considered in the paragraphs which follow

Routine test administration is often administratively desirable In schools and colleges it is most economical to give tests to entering classes, in order to have the data for sectioning, screening, diagnostic, and counseling purposes when and as needed In guidance centers it simplifies scheduling, cuts down expenses, and is a safeguard against failure to obtain and consider basic objective data such as intelligence and interest test results The question then is, can these values be preserved, without creating the mental set criticized by the nondirectivists and most other counselors? The writer is not familiar with any systematic experimentation with this problem, but experience and observation suggest that it may be done by two methods one applicable to academic testing programs and one to guidance centers

In *school or college testing* programs a large number of the students taking tests do so as a routine matter, because they are asked to do so rather than because they want to take them Others are more immediately interested because they have problems of curricular or vocational choice to the solution of which they believe tests will contribute Both of these groups can be helped by a brief explanation of the fact that the testing program is part of the institution s method of obtaining information which may be useful in problems of choice and adjustment, and that the test-obtained data are just one part of the information secured over the years, and added to the student s record It is stressed that the other data, such as the previous school record, grades, extracurricular activities, part-time and summer work experience, and the student s own feelings on these matters of vocational choice and educational adjustment, are of central importance, test data being just one kind of helpful informa-

tion The procedure is like a routine medical examination it may not turn up anything of special significance, but on the other hand some items may help to give a better understanding of a situation This type of explanation will not uproot any strongly imbedded ideas that tests give all the answers, but it will help prevent their taking root and may help pave the way for more individualized structuring when a student comes for counseling

In *guidance centers* to which individuals come for help with problems of vocational and educational adjustment test administration is always preceded by some sort of intake or registration interview This can be and too often is handled as nothing more than a registration procedure, in which the basic data concerning the client are obtained, the presented problem is ascertained, the type of test battery to be given is determined, and an appointment is made for testing But it can also be made an occasion in which the client finds an opportunity to discuss his problem in a permissive atmosphere and to develop an often-needed orientation both to his situation and to the kind of help which can be given by a guidance center If the intake or preliminary interviewer (whether or not he is the final counselor) is nondirective at first and permits the client freely to explore his problem he will generally get better insight into its nature and into the kind of testing and counseling which is needed than if he proceeds at once systematically to take a history, he will establish a relationship in which the client is an active participant, and with this as a foundation he can help the client to understand that the information asked for in taking the history and the data sought in administering tests are simply part of the background material which the counselor uses in getting an orientation to him as a person It can also be stated that some of the background data, such as the type of education received, grades earned, jobs held, and scores made on tests, may at various points in the counseling be facts to which the client and counselor will want to refer and which they may want to discuss Testing may then occupy the same position in the counseling sequence as it now does in most guidance centers, and it may stand out as such administratively, but the client no longer views it as *the* procedure which gives the counselor and himself the answers they seek, instead, he sees it as simply one more data-collecting device, and he begins to understand that data collecting is only one small part of the counseling procedure It then devolves upon the counselor to establish a permissive relationship with the client, carrying on that begun in the early part of the intake interview

This is done by keeping the focus on and the responsibility in the client The counselor may begin the first interview, for example, with a statement that "Mr Doe, the first interviewer, has told me something about you, and of course I have looked over the records, but I think it would be most helpful if you would tell me, in your own words, just what you have been thinking about and with what you think we might help you" The new relationship thus begins as one in which the client is active rather than passive, in which the counselor can accept and reflect feelings, and in which the client uses the superior knowledge and insights of the counselor to develop his own understandings

After routine testing and the re-establishment of a permissive relationship, *test interpretation* may be done at various points during the interviews When the client wants to evaluate himself in comparison with other students or occupational groups, and asks how he stood on some test, the counselor reports the results in actuarial, nonevaluative terms in the manner previously described, permits the client to react to the facts, reflects his feelings, and facilitates further self-evaluation as the client continues to explore the significance of the facts for himself This reporting of test results is often scattered throughout a series of interviews, the data being introduced only as they are relevant and requested by the client More often in practice, but perhaps less desirably, the reporting of test results is done in one session, in which the counselor gives the client a profile of his test results to help him visualize them while he explains their actuarial significance The results should be expressed in percentiles (without I Q 's!) so that relative standing will be easily understood by the client, and the nature of each group with which a comparison is made must be explained as briefly recorded on the test profile Having the data in front of him permits the client to take in both his standing on each test and the nature of the comparison This process in unfamiliar to the client and therefore requires much more of a mental adjustment than the counselor, used to test reports, generally realizes It makes it possible for the counselor either to complete his explanation of the whole battery, allowing the client to come back to and discuss each score, or to stop after each test score has been briefly explained and allow for as thorough an exploration of that datum as the client wants to undertake The writer is not ready to recommend one procedure rather than the other, but suspects that whichever method the client wants to use is best If so, the effective counselor will pause long enough and be permissive enough after each brief explanation for the client to be able to take the

initiative any time he is ready to use it It is, after all, the client who must use the test results, and as their use is an emotionally loaded process it is well to let it be nondirective

Client-determined test administration is perhaps the best term for what Rogerians would call client-centered testing, for the type of routine testing which has already been described is also client-centered in that tests are selected on the basis of relevance to the client and results are used as he needs them to clarify his thinking This procedure has been described by Bordin and Bixler (113) in a way which evokes considerable criticism from some counselors, as they proposed that the client choose his tests himself with little or no help from the counselor It is not necessary, however, to go to this extreme in order for test selection and administration to be client-determined If counseling has been begun nondirectively and the relationship is one in which the client works on problems of his own choice at his own speed, he may, to quote Rogers (640 142), "reach a point where, facing his situation squarely and realistically, he wishes to compare his aptitudes or abilities with those of others for a specific purpose Having formulated some clear goals, he may wish to appraise his own abilities in music, or his aptitude for a medical course, or his general intellectual level " When such desires are expressed the counselor may give the test or tests himself, giving the client the resulting information in the way already described Better still in some cases, the counselor lets the client obtain the information himself by working up the percentiles and test profile together, thus continuing the mutual processes and joint activity of the counseling relationship, the client reacting to test data as they are obtained and the counselor explaining actuarial aspects and reflecting the client's feelings

Combs (166 266) believes that the difficulties in the way of the counselor who attempts to provide information and clarify attitudes are so great as to make reliance on a third person as test administrator and interpreter desirable, with the client then returning to the original counselor for nondirective clarification of feelings and reaching of conclusions As he points out, it takes a good deal of skill to shift from one relationship ("directive" supplier of facts) to the other (nondirective acceptor and reflector of feeling), but then so does any aspect of counseling, and this writer believes that it is a technique which can be learned and used like any other Whether or not it is learned and used depends upon the counselor's theoretical orientation, personal preferences, and work situation In Combs' case all three were in favor of not having one

counselor shift back and forth from more to less directive techniques, in most cases the requirements of the work situation and the desirability of continuity of relationships and work lead the writer to believe that skill in the use of both techniques and in shifting from one to the other is desirable

The Counselor's Moral Responsibilities Breaking Bad News and Sharing Good As a user of psychological tests and as a diagnostician of vocational aptitudes and interests the counselor has available information which may be of crucial importance to the client and of value to society But neither the individual nor society may be aware of the availability and significance of that information, the client may never ask for it, and the counselor may never seek to obtain or to share it, if strictly nondirective procedures are used One might ask whether it is ethical for a counselor to let a high school student work through his attitudes toward going to college and to make college plans without checking up on his mental equipment for going to college Does a counselor who knows that a young man who is planning to enter a skilled trade actually has abilities and interests which might make him a scientist of considerable stature owe it to his client to make him aware of that fact? And does he owe it to society? It is not only attitudes which make for success and satisfaction abilities, opportunities, and awareness also play a part The counselor has an obligation beyond that of assisting the client to assume responsibility for his own action, although that seems to be the sole objective of counseling set up by the nondirective school He also has responsibilities for the detection and optimum use of talent, and for helping some clients to achieve insights into themselves and into society which they might not develop in sufficient time if left to direct the entire course of counseling themselves

The counselor or test interpreter therefore needs to make sure that certain facts basic to the solution of the problems being worked on by the client are secured and considered Intelligence tests may not be asked for in considering the choice of college or of occupational level in such a case the counselor must either be sure from other evidence that the client has the ability to implement his plan or help the client see the need for the obtaining and considering of such evidence Interest inventories may not be requested or interest scores discussed in considering the choice of a field of work, but if the counselor does not see good evidence in the cumulative record or case material that the field being considered is compatible with the client's interests, he owes it to the

client and to society to lead the client to want to obtain such evidence Psychotherapy cures some seemingly physical illnesses and solves some seemingly vocational problems, but it leaves other body ailments and other vocational adjustment problems untouched not to make a diagnosis when diagnosis may be important is as potentially serious an omission in a counselor as in a physician The counselor or psychologist need be no more apologetic about being directive in such instances than is the physician The crucial question has to do with how the counselor brings such evidence to the attention of the client This is primarily a problem of counseling rather than of test interpretation, but as the solution is sometimes sought in test interpretation it should be briefly considered here

To put it in negative terms, a client should not be confronted with an unsuspected low intelligence test score, low musical aptitude scores, or an unfavorable personality inventory score The counselor must instead lead the interview into channels which help the client to explore these characteristics This may be done by getting him to talk about his school grades, his success as a member of a glee club or band, or his relations with fellow-students or fellow-workers Discussion of any of these matters in a permissive atmosphere usually leads the client to examine his aspirations and his disappointments, his strengths and weaknesses (509,627) Reflection of the related feelings encourages the pursuit of these topics During the course of such discussions it is generally easy enough for the counselor to introduce relevant objective data with which the client should be familiar, in fact, the client will often ask for them before the counselor has to take the initiative From then on the process is strictly one of counseling, and beyond the scope of this book

CHAPTER XXII

PREPARING REPORTS OF TEST RESULTS

WRITTEN reports of the results of psychological tests are generally prepared for one or more of the following reasons 1) to provide a permanent record of the interpretations made by the person who counseled the client, 2) to provide an interpretation of the results for the use of other professional workers, 3) to insure that the user of the test results makes a thorough analysis of his data rather than relying on clichés or stereotypes, and, 4) to provide clients or their parents with a record of the interpretations for future reference The first three reasons pertain to the same type of write-up, which may be referred to as the *report to professional workers*, the last may be called the *report to clients* Each of these is taken up in turn in this chapter, from the point of view of purpose, form, and content

Reports to Professional Workers

Depending upon the situation in which he is working, the psychologist who administers tests and reports on test results does so in one of three ways He may simply submit a graphic *profile of test scores* to a counselor or personnel worker in the same organization, accompanied by notes on observations He may make *limited interpretations,* working primarily from the test results, when testing for a colleague to use in working with the client, this user being a counselor, psychiatrist, social or personnel worker Or he may draw on all the case material in making a *full interpretation,* avoiding dependence on the ability of the user to synthesize the test findings with case history material

Profiles of Test Results The most effective way of presenting test results in some guidance centers and business organizations has been found to be the test profile or psychograph This is true when testing is done by psychometrists who have no more skill in interpreting test results than the counselors or executives who use them, and when the

latter have been so well trained in test interpretation that it is uneconomical for the psychologist to write out detailed interpretations which the counselor or personnel worker can make himself In the latter situation the psychologist and the user of the test results generally find that a brief discussion is all the profile ever needs by way of supplementation, or that a few notes at the bottom of the sheet on the client's behavior during testing take care of the subjective factors which the counselor should consider.

The objective of the test profile is, then, to set forth the test results in the simplest and clearest form, so that a trained and experienced user of test results, who can confer with the examiner in case he has questions, can quickly grasp their significance It is also sometimes a useful device for study with a client, serving as a basis for discussion in which the client develops insights by analyzing both the data and his reactions to them He is aided by the counselor's interpretations of the actuarial significance of the tests and reflections of feeling

The principles which govern the development of test profiles and the graphic representation of standing on tests can be outlined as follows 1) tests should be grouped according to type of aptitude or trait measured, 2) when standard batteries are used the test names should be printed on the profile sheet to the left of the grid, but when the tests used vary greatly with the client blank spaces should be left in which test names may be entered, 3) space should be provided after the name of each test for entering data concerning the norm group, 4) another space should be allowed for recording the test score or percentile 5) the grid or graph on which the test results are plotted may be based on either percentiles or standard scores, or may show both, but the users should be conscious of the advantages and disadvantages of both types of scores, 6) some test data are not appropriately represented on the grid together with aptitude and achievement tests and need a special type of presentation on the test profile, 7) space should be provided for supplementary personal data to aid in interpretation, 8) it may be desirable to record observations made during the test sessions to aid in understanding some of the objective scores Each of these principles is taken up in more detail below

1) The grouping of tests according to type of aptitude or trait measured is primarily to facilitate the comparison of test scores which should be approximately the same Although the Minnesota Spatial Relations Test and the Minnesota Paper Form Board measure the same basic

aptitude, the latter test is more affected by general intelligence than the former, for this reason a client's status will not be identical on the two tests, but a study of the differences often helps give a better understanding of the person tested This type of analysis is aided by juxtaposition of the scores In the case of interest inventory scores, which do not lend themselves well to plotting on the grid, parallel listing of the scores of the commonly used inventories is helpful in making comparisons Three profile forms reproduced in this chapter (Figures 9, 10, 11, and 12) illustrate this principle On the form used in the writer's course in vocational testing (Figs 9–10) the aptitude tests are grouped on one page by the side of the grid in the following sequence scholastic aptitudes (3 tests), vocabulary (8 subtests), scientific (6 subtests standardized on technical groups), clerical (2 subtests), manual (6 tests or subtests beginning with gross manual and graded to fine-finger dexterity), spatial relations (4 tests), and mechanical (7 subtests and 3 tests) The classification of tests is not strictly according to traits measured, but also takes into account the nature of the occupations for which the tests have been standardized, a compromise with the imperfections of test construction In the Differential Aptitude Tests (Fig 11) this difficulty is overcome, along with others noted below, by the practice of developing one co-ordinated battery of tests rather than using, as one so often must, a variety of tests from different sources

2) The names printed alongside the grid save time and improve the appearance and readability of the profile if the clients worked with are sufficiently homogeneous in status and objectives for a standard list of tests to be appropriate, selections being normally made from within this list This is true in selection programs in which standard batteries are used, and in specialized guidance centers The Vocational Advisory Service profile (Fig 12) attempts to effect something of a compromise by listing traits measured rather than the test doing the measuring This has the advantage of focusing on psychological characteristics, but makes it necessary to write in the name of the test except when users of the report know that a certain test is routinely used for each trait, as in the case of the spatial and dexterity tests Figure 13 shows a form on which no test names are specified, because of the variety of tests used by the agency, it provides space in which to make note of the names Such flexibility of forms is essential in such an organization

3) Spaces for the entering of data concerning the norm group are essential because most tests have several sets of norms from which the

PSYCHOMETRIC REPORT Sex _____ Age ____

Teachers College, Columbia University

Subject _____ Examiner _____

Aptitudes	Norm	%ile	Letter Ratings: E	D	C	B	A
Scholastic ACE T							
Q							
L							
Calif T							
Q							
L							
Otis I Q							
Voc Mich							
Human Relations							
Commerce							
Government							
Phys Sci							
Biol Sci							
Math							
Arts							
Sports							
Scientific (Engin & Phys Sci)							
Math							
Formulation							
Sci Comp							
Arith Reas							
Verbal Comp							
Mech Comp							
Clerical							
Minn Nos							
Minn Nus							
Manual							
Minn Placing							
Minn Turning							
Purdue Pegboard							
Penn Bimanual							
O'Connor Finger							
O'Connor Tweezer							
Spatial Relations							
Minn P F B Rev							
Minn Spatial							
Crawford Spatial							
Wiggly Block							
Mechanical MacQuarrie							
Trace							
Tap							
Dot							
Copy							
Location							
Blocks							
Pursuit							
O'Rourke							
Minn Assembly							
Bennett							

0 25 50 75 100

FIGURE 9

examiner selects the most appropriate Without these notations the counselor cannot know the significance of the client's standing, as in the case of the Minnesota Clerical Test, on which standings when compared to accountants are radically different from standings when compared to

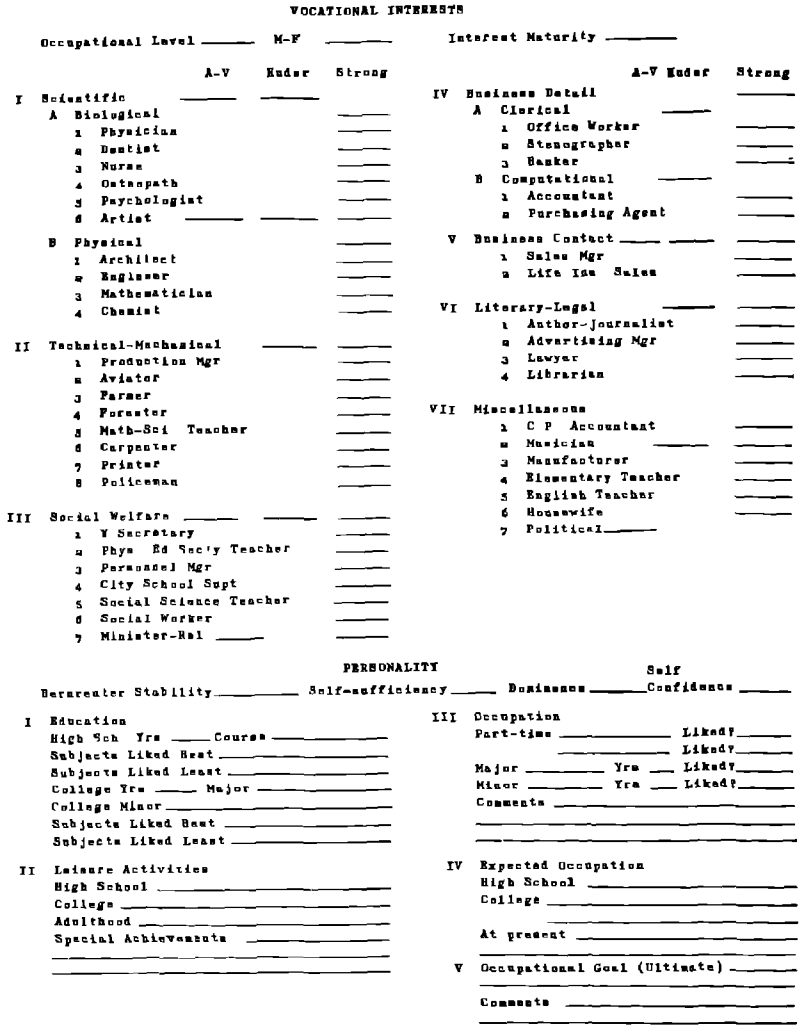


FIGURE 10

INDIVIDUAL
REPORT
FORM

DIFFERENTIAL APTITUDE TESTS

G K. Bennett, H. G Seashore, and A. G Wesman

THE PSYCHOLOGICAL CORPORATION

•New York 18, N Y

NAME SEX AGE (YRS / MOS) GRADE

PLACE OF TESTING FORM NORMS USED DATE OF TESTING

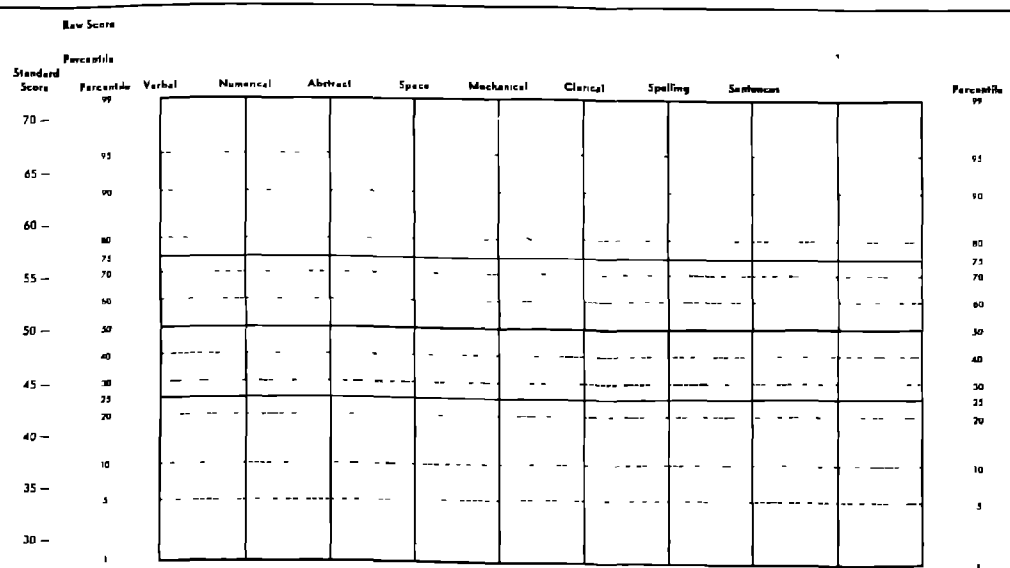


FIGURE 11

VOCATIONAL ADVISORY SERVICE
CONSULTATION SERVICE

PSYCHOLOGICAL PROFILE

Client _____ Counselor _____

Sex- M_____ F_____ Chron. Age _____ Date _____

1	8	31	71	94	99	TEST	SCORE	ERRORS
POOR	BELOW AV.	AVERAGE	ABOVE AV	EXCEPT				
						ACADEMIC		
						Vocabulary		
						Verbal Reasoning		
						Vbl Speed & Alert		
						Wesh -Bel -Verbal		
						CLERICAL		
						Number Checking		
						Name Checking		
						Simple Posting		
						Complex Classif		
						Filing		
						Arithmetic		
						Spelling		
						MANUAL		
						Finger Dexterity		
						Tweezer Dexterity		
						Placing		
						Turning		
						SPATIAL		
						Wesh -Bel -Non V		
						Two Dim. Simple		
						Two Dim Complex		
						Three Dimensional		

Comments.

Entered by _____ Checked by _____

JCA 10 (3-18-46)

FIGURE 12

the general population or even to general clerical workers These notations can be brief, as in Figure 9, for the counselor should know the tests well enough to remember the details if practice calls for no written interpretation by the examiner

4) Space for recording the standard score or percentile obtained when the examinee is compared with the norm group is necessary, both as an aid to plotting the graph and as an aid to using it, for minor errors in plotting and reading graphs are common The numerical entry permits accuracy, just as the graphic entry facilitates grasp of relationships

5) Grids based on percentiles have the advantage of using the familiar and readily understood form of expressing standing on a test in relation to other persons, whereas standard scores are much less commonly known in non-psychologically trained circles But the percentile system has the disadvantage of distorting scores at the extremes, thereby minimizing the important differences in aptitude, while standard scores accurately express these same differences For example, I Q 's of 135 and 190 are both expressed as the 99th percentile, despite the fact that the latter is much further from the mean than the former, making two persons with those scores seem equally intelligent instead of quite different from each other As standard scores are based on distances from the mean, this less-known system reveals instead of hiding this difference However, as it is probably easier to explain this fact to test users than to get them to adopt standard scores, it is probably wiser in practice to use the percentile system and keep its defects in mind When space permits it is therefore wise to provide also for the recording of I Q 's (Fig 9) and standard scores beside the grid

6) Test data which do not lend themselves to presentation on the grid include the results of some interest inventories and some personality measures Scores on Strong's Vocational Interest Blank, being measures of similarity of interests to those of men in various occupations, do not have the same meaning at the higher extremes as do aptitude tests As Strong has pointed out (775 67), there may be no real difference in clerical interests expressed by standard scores of 55 and 65 both persons have interests like those of clerical workers, and the former may perhaps be as representative of clerical workers as the latter Strong therefore rightly recommended use of letter ratings, and these do not lend themselves to plotting on the more refined continuum of percentiles Figure 7 illustrates the profile form used by Strong, combining letter

grades and standard scores Another effective way of organizing such data is shown on the obverse side of another psychometric report form (Fig 10), on which the types of interests measured by the Strong, Kuder and Allport-Vernon inventories are roughly equated and grouped according to types It then becomes possible quickly to scan the entries in order to see in which occupational families high ratings tend to predominate This has the advantage, also, of emphasizing the difference between aptitude and interests, frequently forgotten by clients and by relatively untrained counselors

7) Space for supplementary personal data is something of a safeguard against interpreting test results in a vacuum Most forms call for age and sex at the top of the first page, where they are seen before any test scores The obverse side of one form (Fig 10) provides space for the most important educational, avocational, occupational, and aspirational facts concerning the client These make it possible for the user of test results to check quickly the client's measured interests against his expressed interests and ambitions, and to ascertain whether or not his aptitudes are reflected in experiences appropriate to them The data are too sketchy for complete diagnostic work, but help in case of quick reviews

8) The recording of the observed behavior of the client often cannot and need not be a detailed and tedious task, and therefore generally does not require much space It is important that some evaluative comments can be made on the test profile, however, in special cases Figures 12 and 14 reproduce forms in which space is provided for such notations These are especially desirable in the case of apparatus tests which permit the subjective analysis of the client's approach to problems

Limited Interpretation This is the type of report of test results which should be prepared and used routinely in guidance centers in which psychometric work is done by psychologists, and in which counseling is carried on by vocational counselors who have more knowledge of occupational requirements and of counseling techniques but less of testing It puts the burden of test interpretation where it should be, but leaves the integration of the interpreted test data with the case history material to the skilled counselor who sees the individual and his situation both objectively on paper and dynamically in interviews It is not a worthwhile type of report in situations in which the counselor knows more about tests than the psychometrist, for then the counselor can see more meaning in the test profile than the psychometrist can put into

the write-up Neither is it valuable in situations in which the psychologists knows case study procedures well and the counselor, psychiatrist, social or personnel worker has not had extensive, specific experience in using test results Nor is it likely to prove useful as a report from one agency to another, except when the other agency is an equally well-staffed counseling service In such instances full interpretation on the basis of all available supporting evidence is essential, as borne out by the experience of more than one guidance center which has attempted to render testing services to social agencies without full interpretation the test results generally impress the users as being of little or no practical use

The objective, then, of the limited interpretation of test results is to put into the hands of the counselor a concise, verbal, occupationally-rather than test-oriented statement of the significance of the test scores The counselor then relates them to other data already in his possession or obtained as he works with the client

The principles which apply to the limited interpretation of test results may be stated as 1) the interpretation of each test score first in the light of the appropriate norm group or groups, 2) the relation of that score and percentile to observed behavior in the test situation, 3) the relation of each such interpreted test score to any others which may have bearing on its further interpretation, 4) the modification of this interpretation in the light of any personal data affecting the suitability of the test content or of the norms, 5) the expression of these interpretations in so far as possible first in psychological and then illustratively in broad occupational terms, and, 6) the summarizing of the interpretations to yield a picture of a person and of his occupational potentialities

1) The interpretation of each test score first in the light of the appropriate norm group or groups requires only the verbal statement of what appears in the profile of test results For example "On the 1914 edition of the American Council on Education Psychological Examination he stood at the 97th percentile when compared to freshmen in more than 300 colleges"

2) The relation of this interpretation to observed behavior in the test situation provides an opportunity to mention anything unusual which might have affected the client's performance, such as resistance to taking the tests, undue tension, or concentration and a systematic approach to the task at hand, e g, "He seemed impatient with the

discipline inherent in the test situation and wanted to skip the instructions given before each practice test, but controlled these reactions and worked steadily on the subtests proper."

3) The relation of each test score to others which may have a bearing on its interpretation requires the mental review by the examiner of other data, and the mention in the report of any implications noted. These may consist of such things as the seemingly discrepant scores on two tests of the same aptitude or trait, and the congruence of or lack of agreement between two tests of different types of traits such as aptitude and interest important in the same occupation. An example might be: "The evidence of aptitude for professional or executive endeavor provided by this intelligence test is not supported by the scores of the interest inventories administered."

4) The modification of interpretations such as those given above in the light of personal status affecting the suitability of test content or norms requires a reference to personal history data such as age, sex, education, and cultural background, and consideration of their resemblance to those of the standardization groups. To illustrate: "As the client is now 23 years old, it is probable that his standing when compared to freshmen on the A C E Psychological Examination is somewhat biased in his favor, for it has been demonstrated that scores on this test increase with age during the age range from 18 to 22. Even if his standing on this test is really somewhat lower than it seems, however, the indications are that he is well above the average college freshman in scholastic aptitude. This is borne out by his score on the Wechsler-Bellevue, for which the comparison is with adults in general and shows him to be in the very superior category."

5) The expression of test scores first in psychological and then in educational or vocational terms ensures both the scientific accuracy of the description of the examinee and its meaningfulness to the non-psychologists who often use the results. It provides an educational or occupational sketch of the individual which is more dynamic than a profile of test results. The test summaries and case summaries which follow in this and in the next chapter will serve to illustrate this principle better, but a brief illustration follows. The interpretations which have so far been given might be followed up with: "The conclusion may be drawn that this client has the scholastic aptitude successfully to complete the work of a four-year liberal arts college, although his interest inventory

scores, which remain to be reviewed in detail, suggest that such work may not be exactly to his taste Men with general ability comparable to his tend to gravitate toward professional or managerial work, whether or not they go to college"

6) The summary picture of the person tested, pointing up his educational and vocational potentialities and liabilities, brings together the gist of what has been brough out in connection with the results of specific tests It attempts to integrate these findings into a dynamic picture of psychological characteristics, from which occupational inferences may be drawn by those who know occupations, and of occupational possibilities indicated by the known validities of the tests used The summary of the test report from which the preceding excerpts have been taken attempted to implement this principle in the following way

"In summary, the client appears to be a young man of very superior mental ability, capable of graduating from a good university and rising to positions of considerable responsibility His speed in the perception of clerical detail, particularly numerical symbols, is comparable to that of successful accountants His superior ability to judge shapes and sizes and mentally to manipulate them is indicative of promise in technical and artistic occupations In ability to perceive and analyze the effects of physical forces and the operation of mechanical principles he does not compare well with engineers or skilled artisans, although he compares favorably with the general population His interests are not highly developed in any area, although they resemble somewhat those of men who are engaged in business occupations involving contact with other persons and the management of enterprises Experience has shown that many young men with abilities and interests such as this client's enter business and find their way into executive positions"

Common errors in the preparation of reports of test results generally consist of violations of the above principles Psychometrists tend to write in terms of test scores or percentiles rather than in terms of aptitudes and traits, psychologists without extensive contact with business and industry sometimes find it difficult to translate psychological characteristics into occupational behavior, and those who are not well grounded in both psychology and in occupations tend to overwork the brief and somewhat stereotyped interpretive phrases of the test manuals The result is test-centered, and the real significance and value of testing is lost Perhaps an illustration of poor reporting, accompanied by an

Test Record and Profile Chart

Name _____ C-number _____

Test No.	Test Name and Form (or Subscore)	Norm Group	Raw Score	T-Score or Standard Score	Percentile Rank
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					

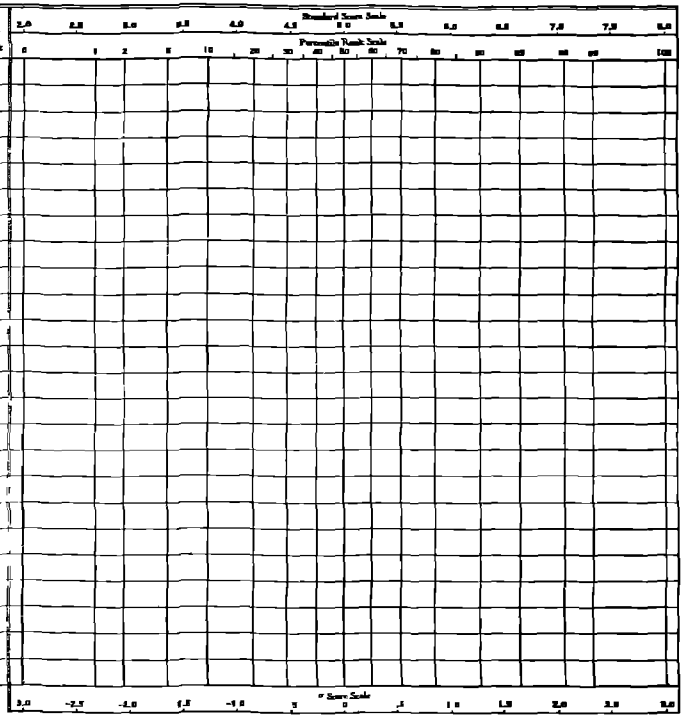


FIGURE 13

Test No.	Date of Test	Examiner	Examiner's Title	Institution	Examiner's Observations and Comments
1					(Refer by number to any particular test, include any necessary additional explanations of special items such as norm groups)
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					

improved version of the same report, will help to illustrate some of these points To facilitate comparison they are reproduced in paired, original and revised, lines when changes seem desirable

In summary, the client's very { high scores on the Paper Form Board, the / superior ability to visualize space relations,

performance subtests of the Wechsler Bellevue Scale, and the Meier Art to think in non-verbal abstractions, and to judge the quality of form and

Judgment Test, / composition, } indicate a great deal of potential ability in art { work or in / and in re-

work involving spatial as well as aesthetic judgment / lated types of work, } such as layout or pro-

duction work in an advertising agency The low score on the mechanical comprehension test suggests artistic rather than technical outlets for her

spatial ability / ability to work in spatial arrangements } The average { clerical scores and the / speed of perception of

48th percentile "effectiveness of expression and accuracy / clerical symbols and average clarity } of written expres-

sion" suggest that she will not be handicapped in these activities { should they / but that she

be involved in her work / would do well not to specialize in business detail or linguistic work } Her

intelligence, which is { better than that of 97 percent of the general popula / very superior suggests ability to rise high in any work

tion, should ensure ability to succeed in any area using her other aptitudes utilizing her other aptitudes and providing outlets for her interests

Briefly stated, the changes in the above summary were intended to produce a description of the client s abilities, interests, and occupational promise, rather than a summary of her test scores Perhaps this type of report can best be made clear, after the outlining of principles which has just preceded, by reproducing *in toto* a report written for use by a trained counselor working in the same agency and for possible sending to a similar college counseling bureau Such a report follows, with data slightly changed and identity disguised

REPORT OF TEST RESULTS JOHN F ATKINSON
(Limited Interpretation)

John Atkinson, a high school graduate, 19 years of age, was given two tests of scholastic aptitude On the Otis-Self-Administering Test of Mental Ability he was at the 50th percentile when compared to college students, which would suggest that his chances of competing with college students and completing college work in an average college are reasonably good On the 1944 Edition of the American Council of Education Psychological Examination he was at the 27th percentile when compared to college freshmen His linguistic score was at the 33rd percentile and his quantitative score at the 26th percentile

As the A C E test is a somewhat longer and more appropriate instrument, this suggests that John, while able to compete with college students, is likely to find himself in the lower third of the student body and will therefore find it necessary to apply himself more effectively than the average college student in order to achieve satisfactory results

Several tests of special aptitudes which are important in engineering and scientific occupations were administered On the Engineering and Physical Science Aptitude Tests his mathematical score was at the 4th percentile, his physical science comprehension score at the 12th percentile, and his mechanical comprehension score at the 39th, on the other hand his arithmetic score was at the 56th percentile, his formulation score at the 70th percentile, and verbal comprehension score at the 74th, compared to men students in non-collegiate technical courses These scores suggest that the client has more aptitude for work of a verbal nature than for mechanical or mathematical work The relatively low standing in these latter areas is confirmed by the Minnesota Spatial Relations Test on which he scored at the 30th percentile when compared to engineering freshmen On the O'Connor Wiggly Block his letter rating was C, which points up even more the lack of special aptitude in spatial visualization

One test of clerical aptitude was administered The Psychological Corporation General Clerical Test The total score on this test was at the 29th percentile when compared to clerical workers, the lowest part score being that for clerical speed and accuracy at the 19th percentile and the highest part being verbal facility at the 49th percentile These results fit in with the data indicating greater verbal facility than nu-

merical or spatial, but do not indicate special qualifications for clerical work At the same time the scores are high enough to indicate reasonable chance of success in such employment if other things are favorable

Three measures of interest were obtained On Strong's Vocational Interest Blank John revealed a pattern of interests most similar to those of engineers, chemists, and other men engaged successfully in physical science occupations His interests are also similar to those of teachers of high school science, and to those of production managers and others engaged in semi-technical industrial work, they also resemble those of musicians According to Strong's Blank his interests do not greatly resemble those of men in artistic work, biological science, social welfare, business contact, or literary and legal occupations There is some sign of interest in business detail occupations, including office worker and purchasing agent

The results of the Kuder Preference Record did not agree very well with the results of the Strong Blank, although scores on the two tests tend to confirm each other in most cases According to the Kuder, John's interests are strongest in artistic, musical, social welfare, and mechanical activities The moderately high interest in music and in technical work indicates some agreement with Strong's Blank, but there is a real discrepancy between the two tests on artistic and social welfare interests

The Allport-Vernon Study of Values throws some light on this matter by showing a fairly high social welfare score and an average theoretical or scientific score, but confuses the issue by revealing a strong interest in material welfare, such as normally characterizes men in business contact occupations What seems fairly clear from these interest test results is that John does have interests comparable to those of men who are successful in managerial work in industry, the picture is not clear cut with reference to other interests

In summary, it would seem that John Atkinson is a young man of fair college aptitude, greater linguistic than scientific aptitude, and interests which most clearly resemble those of men engaged in managerial and supervisory positions in industry His prospects for success in a four-year engineering or liberal arts college do not seem especially good, although such students do graduate from the less competitive colleges On the other hand it seems likely that he would succeed in an industrial engineering course or in a business course aiming at administrative work, taken in an institution in which the competition is not too great

Full Interpretation As previously stated, this type of report of test

results is most likely to prove valuable when reports are prepared by well-trained and experienced vocational psychologists, particularly if they are to be used by counselors, psychiatrists, and social or personnel workers who have not been trained in the use of test results In such instances the psychologist shares the other workers' ability to make a case study, and has, in addition, the knowledge of tests, and of their occupational significance, not possessed by his colleagues The effective use of talents then calls for full interpretation by the psychologist, although their application in counseling or in selection and promotion may be made by other specialists, depending upon the nature of the situation and of the case The counselor may need the data in connection with educational and vocational planning, the psychiatrist in connection with therapy which calls for the most effective use of his patient's abilities and interests, the social worker as an indication of the types of vocational rehabilitation which may be effective, and the personnel worker as an aid to making decisions concerning employment and advancement And the psychologist who functions as a clinical counselor will also find that the preparation of such a report is one of the most effective methods of forcing himself fully to explore the significance of test results and personal history data

The objective of full interpretation is to tease the maximum of meaning from the test results by synthesizing them with other case-history material, at the same time using one type of data as a check on the other in preparing an accurate and vivid description of the person being studied

The principles guiding the preparation of full interpretations of test results are the same as the six governing limited interpretations, with the addition of one more which follows after the fourth in the list given on page 574 This principle specifies the necessity of viewing the test data in the light of related case history material which it may confirm, contradict, or illuminate, or by which it may be confirmed, contradicted, or illuminated

Viewing test results in the light of other case material requires that the interpreter be trained and experienced in case-history taking and in the occupational and clinical significance of personal, socio-economic, educational, and avocational data The process involves the examination of intelligence test results in the light of educational attainment, the comparison of measured interests with interests as manifested in school subjects, leisure-time activities, and previous occupational experience,

and the evaluation of special aptitude test scores in the light of accomplishments in related activities To illustrate "The Co-operative General Mathematics Test for High School Classes was also administered The client had three and one-half years of high school mathematics, followed by college training in accounting, a master's degree in business education, and three years as a junior accountant, in all of which he worked with figures When compared with the four-year norm group he is at the 4th percentile, while with the three-year group he is at the 32nd This low score is congruent with his low quantitative score on the scholastic aptitude test, his own statement that he feels weak in mathematics, and the fact that he failed the teaching examination in mathematics The picture is clearly one of weakness in the mathematical area, although whether or not this weakness is the result of lack of aptitude, emotional maladjustment, or a combination of the two is not brought out by these data "

A complete report of test results in which full interpretation has been attempted is reproduced below, as the best way of conveying an idea of the principles and method

REPORT OF TEST RESULTS JAMES L FRANK
(Full Interpretation)

James is an eighteen year old high-school senior, who came for help in the choice of a career Specifically, he wanted to know whether or not he should go into engineering His father owns a manufacturing plant, he is interested in having the boy go to college and feels that he may be better qualified for administrative than technical work His father thinks that industrial management would probably be a good field The boy worked in his father's plant during one summer but got a different job last summer, working as a bell hop in a resort He preferred to go out on his own rather than into a job already made for him He, too, feels that he is really more interested in administrative than in technical work

James was given the American Council on Education Psychological Examination, two different forms a week apart On the first, he scored at the 74th percentile compared to entering College Freshman, and on the second, at the 76th percentile On both tests, his linguistic score was distinctly higher than his quantitative, which suggests that his and his father's hunch that James is not as strong in the technical field as in others has a foundation in fact

James also took the Engineering and Physical Science Aptitude Test His scores, when compared to recent high school graduates applying

for non-collegiate technical training, were in the bottom decile for arithmetic reasoning, in the fourth decile for mathematics and formulation, almost at the 75th percentile in physical science comprehension, and in the top decile for verbal comprehension and mechanical comprehension These results suggest that, while James is weak in mathematical ability, he does have a rather high degree not only of verbal but also of mechanical aptitude

The Minnesota Spatial Relations Test was given, the letter grade being B— As the norm group is the general population, it seems legitimate to conclude that James does have a relatively low degree of ability to visualize spatial relations

The Purdue Pegboard was administered, James being near the 99th percentile on all part scores and total scores when compared with college men

The Strong Vocational Interest Blank shows no primary interest patterns and no A ratings The greatest concentrations of interests are in the physical sciences, technical and social welfare fields James rates a B— as engineer and chemist, C as mathematician, B+ as mathematics-science teacher, B as production manager, B as personnel manager, B— as Y M C A physical director and social science teacher, C+ as Y M C A secretary, C as school superintendent In the other fields, James's scores are scattered B—'s and C's His interests are like those of many young men who go into business in that they are relatively undifferentiated But he is somewhat stronger in the practical side of technical work and in the fields of human relations and personal contact

The Bernreuter Personality Inventory indicates that James is an emotionally stable, somewhat dependent, extroverted, rather dominant, self-confident and quite sociable young man.

The rest of the results appear to agree quite well with the interview material James's extra-curricular and leisure-time activities are primarily social, and indicate not only interest but considerable skill in dealing with people His grades in mathematics are poor but are acceptable in other subjects His general ability is better than that of the average college freshman, but he lacks some of the special aptitudes required for success in a technical curriculum He has certain other aptitudes which would be assets to him, particularly his mechanical comprehension and his verbal ability James would probably do well in a position such as his father's, in which facility in understanding mechanical processes is necessary, and in which finding some engineering activities con-

genial would help His personality characteristics would also be an asset in the supervision and contact side of industrial management

In choosing a college, James would probably do well to select one in which he can follow a business administration or industrial engineering major It would be helpful if summer employment could be obtained in an industry (other than his father's) rather than in a field unrelated to his educational and vocational objectives This would permit him to try himself out and get the experience of making his own way It would make it emotionally easier for him to return to work in his father's plant within a relatively brief time after the completion of his education if that seemed desirable

Reports to Clients

The problem of preparing reports of test results for clients who have been counseled is a vexatious one Ideally the counseling of which testing and test interpretation are a part should have been so conducted that the client (and his parents, if they are involved) has integrated the test results into his own thinking He then has insights into their significance which match his understanding of his school record and his vocational experiences and views them in very much the same light Just as he does not think it necessary to have a written record of all of his jobs and of his performance on them, so he should not need to want a written report of the results of his tests When he does it generally means that they are thought of as a crutch of some sort

In a world in which there are crippled people crutches are sometimes desirable They are to be frowned upon only when they contribute to keeping a person partially crippled longer than he need be The fact that clients often want written reports indicates a need for a crutch in some cases When such requests are at all frequent the counselor should examine his practices, in order to find out why his clients seem to feel the need of something tangible to lean on

In the writer's experience, to which appeal is made only because no studies seem to have been made of this question, clients have wanted copies of test reports only, 1) when testing has been overemphasized, 2) when the discussion of test results was not successfully integrated with counseling, and 3) when the client's own insecurity led him to believe that he could use a report of test results to sell himself to a potential employer more successfully than he could on the basis of his experience education, and conduct in the employment interview Before discussing

the form and content of such written reports to clients as are prepared, it may therefore be wise to deal briefly with methods of handling these problems

Methods of Handling Client Requests for Test Reports Methods of avoiding overemphasis on testing and of successfully integrating test results with counseling have been discussed in some detail in Chapter 21, and need not be gone into here But when a client requests a written report of test results which have already been discussed such preventive techniques are no longer usable In the experience of the writer and the counselors whose work he has supervised it has generally been effective to ask the client how he expects to use the report When emotionally insecure clients reply that they will show it to prospective employers as evidence of their qualifications for a job, the counselor asks the client to put himself in the place of the employer He is to consider how he would react to an applicant who applied for a position and produced a sheet of test results to prove his qualifications This generally brings about a realization of the artificiality of the technique and of the fact that most employers still judge in terms of other types of evidence If this realization does not come at once it can be helped by asking how often employers have asked the client if he had any test results to show his qualifications, or had requested that he obtain such The client then generally recognizes that if employers were inclined to depend upon or to be much impressed by the results of tests given by organizations other than their own they would ask for them more often

The client may state that he would like to have the test report for incidental use when talking with employers, that they are tending to become test conscious and might be impressed by the fact that the client had taken the trouble to study himself so thoroughly before applying for a job The counselor can use this as an opening for discussing job hunting techniques, introducing the client to books such as the Edlunds' (236) This helps the client to see that he can best demonstrate the care with which he has gone about seeking employment by an intelligent understanding of the company for which he wants to work and of the ways in which he can serve it, and that merely having a report of tests (of someone else s insights rather than of his own) is likely to be of little value The counselor s suggestion that any employer interested in obtaining a report of the client's test results might write for such a report with the client s permission generally appeals at this point as a more effective way of putting the test results to work than taking a report with him

In certain other cases it may be desirable to send copies of test reports to parents, school principals, or potential employers who may not be well qualified to interpret test results and whose discretion in their use cannot be taken for granted. It should be recognized that the sending of reports to parents is at best a compromise with an undesirable situation, and that if reports from the counselor to parents are necessary they should really be made as a part of counseling. School and industrial users of test results are in a different category, but even in their case it would be preferable from both client's and school's or industry's point of view if the psychologist were to make his report to the principal or employer on the basis not only of familiarity with the client but also with the school or work situation, helping the recipient of the report to integrate client-data with situational data. Because written reports are often the best possible compromise in such situations, methods of interpreting the results in writing are discussed in the next paragraphs.

The objectives of the report to parents or other laymen are the increasing of their understanding of the abilities, interests and personality of the client. As the recipients of these reports are not thoroughly trained in psychology, testing, or counseling, the report aims to describe these not in psychological terms but in terms of their educational and vocational implications.

The principles governing the writing of reports to laymen may therefore be formulated as follows: 1) test names, test scores, and most psychological trait and aptitude terminology should be avoided in favor of descriptions of probable educational and vocational performance, 2) these descriptions should be phrased in broad terms and illustrated with typical concrete examples, and 3) a brief summary giving a dynamic picture of the individual should bring together the interpretations of the more specific aspects of probable behavior. Each of these principles is taken up briefly below.

1) The substitution of descriptions of probable behavior in educational and vocational situations for psychological terminology requires the making of statements concerning probable success in college, technical institutes, and appropriate types of occupations, rather than the description of mental status. It involves comparisons of the interests of the client with those men in occupations which he is or perhaps should be considering, rather than in terms of letter ratings or percentiles. These varied actuarial comparisons are both more meaningful and less traumatic than descriptions of ability levels or personality traits would be. It was written of one

high school sophomore with an intelligence quotient of 90 and no A's or B's on Strong's Vocational Interest Blank "His chances of doing good work in a college preparatory course are slight, but it is probable that he could complete the graduation requirements of the general, commercial, or trade curriculum His general ability is equal to that of many men who have succeeded in skilled trades, such as machinist, printer, or plumber, but he might have difficulty competing in the more demanding aspects of technical work such as mathematics and blue print reading, he could probably compete successfully with routine clerical workers such as stock clerks and general clerks, but would not be likely to rise to a position of responsibility in office work, as a machine operator or assembly worker in a factory he would be competing with men of his own ability level, and could, other things being equal, rise to a position of leadership as a foreman or supervisor His interests do not resemble those of men engaged in engineering, business, or skilled occupations, suggesting that he may find most satisfaction in work which does not require a great deal of specialized information, instead, he may find his satisfactions more in his contacts with other people or in outside activities Many men with interests and abilities like his are more interested in a job with regular hours, steady pay, opportunities to make friends with other people, and time off in which to indulge in special interests such as sports association with men friends, and reading newspapers and magazines, than in the exact nature of the work they do

2) The couching of such statements in terms sufficiently broad to avoid the appearance of prescription but concrete enough to be meaningful, and the use of specific examples which are illustrative rather than limiting is perhaps the most difficult part of writing reports of this type They require considerable knowledge of occupations and of the world of work on the part of the person writing the report Considerable help can be obtained from the literature, e g through the use of occupational norms such as those published for intelligence tests after both World Wars (see Ch 6) and for various types of tests by projects such as the Minnesota Employment Stabilization Research Institute (589), and through familiarity with studies of workers such as those resulting from the Western Electric experiments (637) and the Yankee City studies (909) The illustration in the preceding paragraph should serve for this principle also

3) The summary statement at the close of the report serves to pull together the gist of what has been said before and to avoid the overemphasis of isolated statements which might happen to impress the reader

Of the boy partly described earlier in this discussion it might be said, for example, "In summary, John is a boy who should be able to complete a high school education in the general, commercial, or trade curriculum if he so desires His abilities and interests suggest that he is most likely to find success and satisfaction in the middle range of occupations, and, in that group, most probably in the less competitive general office or factory jobs It seems probable that he will derive more satisfaction from employment which permits him to have interesting friendships and recreational outlets than from some one type of work requiring special preparation over a long period of time John s aim might well be the ability to shift readily from one type of factory operation to another, skill in getting along with people, and knowledge of a variety of industrial processes which make a valued employee in his own company and a very employable applicant in the eyes of other concerns "

CHAPTER XXIII
ILLUSTRATIVE CASES
DATA AND COUNSELING

THERE are so many different types of tests, so many tests of each type, and so many studies of the validity of some of these tests, that it is difficult in books on testing to find adequate space for discussion of the ultimate purpose of testing achieving insight (by the client) and an understanding of a person (by the counselor) An effort has been made to deal systematically with this topic in Chapter 20, and to treat problems of reporting test results to clients in Chapter 21, it still remains, however, to describe the diagnosis and counseling of a number of individuals, and to report their subsequent vocational adjustments, in order to show how the test data were used and how well the deductions from them foreshadowed subsequent developments

Opportunity should also be provided for the student of testing to put to work the insights which he has developed from the contents of this book and from his own experience, by presenting the test data and essential case material in such a manner as to permit him to make his own appraisals before reading those made by the counselors who actually handled the cases The reader may also want to attempt to predict the subsequent educational and occupational histories of the boys and girls, men and women, described by the case summaries It should prove instructive to see how well the reader's and the counselors' insights corresponded with what actually took place From such comparisons one gains new insights into the meanings of test scores, the interplay between various types of personal characteristics, and the interaction of personal characteristics and social environment

The seven cases described in this and the following chapter were tested and counseled by the writer, his associates, and his students in a number of different places at various times during the past 15 years These places were the Cleveland Guidance Service of the National Youth Administration in Ohio, the Guidance Service operated by Clark Uni-

versity in co-operation with a number of high schools in central Massachusetts, the Psychological Services Branch of the AAF Regional and Convalescent Hospital in Miami Beach and Coral Gables, Florida, and the Guidance Laboratory of Teachers College Columbia University For ethical reasons, even the place of work with any individual case and the identity of the counselor (sometimes the writer but sometimes an associate or student) as well as all more personal identifying data such as names and institutions, are disguised

The method of presentation requires a word of explanation in order that the reader may obtain the maximum desired benefit from the material The case histories are divided into three sections 1) case summaries and test profiles, 2) counselors' interpretations and the immediate outcomes of counseling and 3) follow up reports Within each of these three sections the cases are presented in the same order, beginning with boys and girls first counseled as high school students and closing with experienced men and women who came for counseling because they were considering changing occupations Thus Tom Stiles' background data and test profile are presented first followed by those for Marjorie Miller Ralph Sheridan, etc Then the sequence begins again, this time for giving the counselors interpretations and the plans, if any, made by the client (It may be of interest that this material was written up for publication *before* the follow up data were obtained, to avoid contamination by hindsight) The sequence then begins over again for the last time, to show the current status of each of the cases in turn and to consider the validity of the counselors' appraisals It is suggested that readers interested in obtaining the maximum possible value from this chapter make their own diagnosis (and prognosis if so inclined) after reading the background material and studying the test profile of each case, add anything they wish to this after reading the account of the counselors work, and then compare their notes with the follow-up data in the next chapter as these are read for the first time

Background Data and Test Profiles

The Case of Thomas Stiles When is an Engineer an Engineer?

Tom was 17 years old in good health of average height and weight, a high school senior when he came to the counselor He was enrolled in the academic course, in which he liked the work in mathematics and science better than anything else, and cared least for English and history His

leisure-time activities consisted largely of spectator sports, he liked also to read popular scientific and adventure story magazines As a younger boy he had done odd jobs at home, and since then had had part-time and summer jobs working as a helper on a truck, operating machines in a shoe factory, helping in a garage, and working in a machine shop Some of these jobs had been for no pay, others, the more recent, had been paid work

The student's father was an operative in a shoe factory, the mother kept house, and several siblings, all younger than Tom, were still in school

Tom stated that he was interested in machines, having lived among various types of machinery all his life, his junior high school ambition

FIGURE 15

GRADES THOMAS STILES

	Marks			
Subjects	*9th Grade*	*10th Grade*	*11th Grade*	*12th Grade (1st sem.)*
English	C	D	E	III-C IV-C
Latin I		C		
Civics	B			
World History	B			
Prob of Democracy			C	
U S History				C
Algebra I		B		
Plane Geometry		B		
Review Math			C	
Math (Gen.)	B			
Solid Geometry				C
Physics			D	
General Chem				C
Phys Education		D		

had been to be a diesel engineer or marine engineer, an ambition which had broadened to include work with almost any type of engine steam, diesel, or airplane, especially the last named type, as "it is the coming field" He thought he would like engineering training, but was not certain of his choice Asked what he would like to be ten years hence he replied "Foreman or superintendent in an airplane factory"

The cumulative record in the school office showed that Tom's high school work was mediocre As shown in the accompanying chart, he had failed junior English, did poorly in physics had made only C's in mathematics after the 10th grade, and was doing no better in chemistry His I Q on the Henmon-Nelson Test of Mental Ability, administered at the beginning of his junior year and recorded on the school record, was 106

The profile of test results obtained by the counselor during the first semester of Tom's senior year in high school is shown in Figure 16

Tom's questions were "Should I go into engineering? I am interested in engines Should I continue my education in order to prepare for such work? What about engineering college?

FIGURE 16

TEST PROFILE THOMAS STILES

		Norms	*%ile*
Scholastic Aptitude	A C E Psych Exam	Coll Fresh	14
	Otis S A I Q 101	Students	19
Reading	Nelson Denny Vocab	Fresh	1
	Paragraph	"	30
Achievement	Coop Social Studies	"	11
	Coop Mathematics	"	74
	Coop Natural Sciences	"	5
Clerical Aptitude	Minn Clerical Numbers	Gen'l Clerks	6
	Names	"	6
Mechanical Ability	O Rourke Mechanical Aptitude	Men in Gen'l	67
Spatial Relations	Minn Paper Form Board, Rev	Coll Fresh	65
Personality	Calif Pers Self	"	35
	Social	"	40
	Total	"	40

VOCATIONAL INTERESTS

1 Biological Sciences	B+	4 Social Sciences	C
2 Physical Sciences	A	5 Business Detail	B−
3 Technical		6 Business Contact	B
Carpenter	A	7 Literary	B
Policeman	A		
Farmer	B+		

Exercise 1

a) Prepare a written analysis of the test results of this case as though for transmittal to her grade adviser Use the sample on page 582 as a model Do this before reading further and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client, including your approach to counseling the client in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

Marjorie Miller A Case of College and Choice of a Scientific Field

Marjorie was 16 years old when counseling began She was then an academic high school senior, in excellent health, of average height and

weight, very good looking, friendly, and mature in manner She reported that she liked chemistry, languages, and history best, and had no special dislikes in school Her leisure-time activities consisted of photography, dramatic club, work on the school paper, scouts (Mariner, in charge of younger troop), participant sports, dancing, and painting, in this last connection, she had entered some of her work in local exhibits Her reading consisted largely of school-required books Her part-time and summer work experience consisted of selling Christmas cards and working in a gift shop

This pupil's father was employed as an executive by an insurance company, the mother was a housewife, there were no brothers or sisters

Marjorie's plans were to go to a liberal arts college, but in doing so

FIGURE 17

GRADES MARJORIE MILLER

Marks

Subjects	*9th Grade*	*10th Grade*	*11th Grade*	*12th Grade to date*
English	91	88	90	92
French			93	90
Latin	94	94	94	
History	Ancient 94			U S 90
Mathematics	Algebra 88	Geometry 87	Rev Geom 91 " Math 85	Sr Alg 85
Biology		94		
Chemistry			90	
Art	91	91		
Typewriting			80	
Physical Education	88	91	95	

she wanted to 'specialize in some definite subject so as to be ready to work" after graduation She was considering two nearby colleges of good but not outstanding reputation neither of which was actually a liberal arts college but both of which had good professional and business curricula Her occupational preferences were chemical research ("I think I would like the work"), dietetics ("I like the subject") or the teaching of chemistry in high school or nursing school ("If I had to teach, I would want it to be chemistry'), but she was undecided as to her actual choice She had previously thought of art, surgery, medical laboratory work, and tea room management, in that order, beginning in the last years of grade school Ten years hence she wished to be connected in some way with science or medicine "

The high school record showed that Marjorie had done uniformly superior work Her grades were close to or above 90 in all subjects

except for 85 in Review Mathematics and Senior Algebra and 80 in Typewriting. The only patterning revealed is perhaps slightly less strength in mathematics than in the more verbal subjects. She said "I would rather spend time on chemistry than on any other subject. I am interested in math but find it rather hard. I have never taken social studies [this despite a current course in history] but I'm sure I'd like them." The principal described Marjorie as "a brilliant girl with unusual ambition, with many interests, particularly in science." Marjorie's test profile, obtained during the first semester of 12th grade, is reproduced in Figure 18.

The statement of the problem as seen by Marjorie was to choose be-

FIGURE 18

TEST PROFILE MARJORIE MILLER

		Norms	*%ile*
Scholastic Aptitude	A C E Psych Exam	Coll Fresh	72
	Otis S A I Q 124	Students	87
Reading	Nelson Denny Vocab	Fresh	80
	Paragraph	"	96
Achievement	Coop Social Studies	"	50
	Coop Mathematics	"	90
	Coop Natural Sciences	"	70
Aptitudes	Minn Clerical Numbers	Clerical Workers, 1	57
	Names	"	34
	Minn Paper Form Bd Rev	Coll Fresh	70
Personality	Calif Self-Adjustment	"	80
	Social Adjustment	"	40

VOCATIONAL INTERESTS

	Strong	*Allport Vernon %ile*		*Strong*	*Allport-Vernon %ile*
Biological Sciences		90	Domestic		
Physician	B+		Housewife	A	
Dentist	B				
Nurse	A		Business Contact		
Artist	B−	7	Life Insurance		
Math-Sci Teacher	B−		Saleswoman	C'	67
Social Sciences		35	Literary		
Y W Secy	B		Author-Journalist	C+	
Social Wkr	B+		English Teacher	C	
Social Sci Teacher	C		Lawyer	C	
			Librarian	C	
Business Detail			Femininity	72nd %ile	
Office Clerk	C+		Political (Prestige)		47
Stenographer	C+		Religious		92

tween dietetics and chemical research, to decide what kind of training to get and where, and to find out more about the kinds of jobs that might be available to her after completing college

Exercise 2

a) Prepare a written analysis of the test results of this case as though for transmittal to his grade adviser Use the sample on page 582 as a model Do this before reading further and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client including your approach to counseling the client in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

Ralph Sheridan A Case of College and Finances

Ralph was a 17 year old high school senior when tested, a boy of some thing above average height heavily built, in good health, pleasant to look at and to talk with He was enrolled in the college preparatory course,

FIGURE 19

GRADES RALPH SHERIDAN

	Marks			
Subjects	*9th Grade*	*10th Grade*	*11th Grade*	*12th Grade to date*
History	—		90	
English	—	92	91	91
French		86	86	
Latin	—	80	83	
Geometry		88		
Trigonometry				88
Rev Algebra				88
Chemistry			85	
Physics				87
Bookkeeping		91		

and gave English and mathematics as his favorite subjects, and foreign languages, especially Latin as his least liked His leisure time activities consisted of hunting, fishing, trapping and other solitary or small group outdoor activities he was active in debating at school, his reading consisted largely of adventure and historical fiction During spare time and summers he had worked as a berry picker (when younger) in a general store, and on a farm

The Sheridan family consisted of Ralph's father, who owned and

operated a combination general store and service station, his mother, housewife, and younger brothers and sisters

Ralph expected to continue his education after graduation, had definitely decided to be a civil engineer but did not know which college to go to or how to finance it His second and third preferences consisted of lumberjack and farmer, these latter because he 'liked the work', engineering was chosen because 'you can make money and the work is

FIGURE 20

TEST PROFILE RALPH SHERIDAN

		Norms	*%ile*
Scholastic Aptitude	A C E Psych Exam	Coll Fresh	94
	Otis S A I Q 125	Students	89
Reading	Nelson Denny Vocab	Fresh	99
	Paragraph	"	99
Achievement	Coop Social Studies	"	95
	Coop Mathematics	"	82
	Coop Natural Sciences	"	77
Clerical Aptitude	Minn Clerical Numbers	Gen'l Clerks	26
	Names	"	71
Mechanical Ability	O'Rourke Mechanical Aptitude	Men in Gen I	86
Spatial Relations	Minn Paper Form Board, Rev	Coll Fresh	50
Personality	Calif Self-Adjustment	"	30
	Social "	"	20

VOCATIONAL INTERESTS

Biological Sciences	C+	Social Sciences	C+
Physical Sciences	B	Business Detail	B
Technical		Business Contact	A
Carpenter	C+	Literary	B
Policeman	B−		
Farmer	B+		

rather pleasant ' Ten years hence he wanted to be "alive and rich," the first part of which ambition seemed understandable during the Battle of Britain

The cumulative record showed that Ralph's high school work had been of college caliber as all but his Latin grades had been 85 or above, and even they had been above 80 Pattern analysis showed that his grades in verbal subjects were slightly, but perhaps not significantly, higher than in quantitative subjects

The problem, as expressed by Ralph was to choose an engineering college and to find a way of financing it

Exercise 3

a) Prepare written analysis of the test results of this case, as though for transmittal to his grade adviser Use the sample on page 582 as a model Do this before reading further and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client, including your approach to counseling the client in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

Paul Manuelli: A Problem of Choosing a College Major

Paul was a 17 year old high school senior taking college preparatory work and enjoying mathematics and science most while caring least for history, etc ' He was a very tall, well built individual with excellent health, a pleasant appearance, and agreeable manner His leisure time activities consisted largely of participant sports (in which he excelled), parties, and dancing Part time and vacation work occupied a good deal of his time, and consisted of cooking, soda jerking, and mowing lawns, while in junior high school he had worked as a caddy, and had raised vegetables and sold them

FIGURE 21

GRADES PAUL MANUELLI

	Marks			
Subjects	*9th Grade*	*10th Grade*	*11th Grade*	*12th Grade*
Drawing	90			
Ancient History	90			
U S History & Gov t				90
English	90	90	90	90
Public Expression			95	95
Latin	90	90		
French		85	95	
Algebra	90		90	
Geometry		90		(Solid) 85
College Chemistry			85	
College Physics				85
Physical Education	85			

The Manuelli family included the father, employed as an operative in a local factory the mother, a housewife, an older brother who worked in a factory like the father, an older sister then in training as a nurse, and two younger sisters

Paul was not sure about continuing his education beyond high school but hoped to be able to go to engineering school He had saved his summer earnings, but needed more money to help finance his education

He was considering engineering and law as occupations, the former because he liked the related school subjects, the latter because he enjoyed debating and public speaking He was thinking of West Point as a means of combining his engineering interest with the possibility of war, which was then going on in Europe

FIGURE 22

TEST PROFILE PAUL MANUELLI

		Norms	*%ile*
Scholastic Aptitude	A C E Psych Exam	Coll Fresh	87
	Otis S A I Q 123	Student	84
Reading	Nelson Deany Vocab	Fresh	97
	Paragraph	"	96
Achievement	Coop Social Studies	"	73
	Coop Mathematics	"	94
	Coop Natural Sciences	"	24
Clerical Aptitude	Minn Clerical Numbers	Gen l Clerks	26
	Names		53
Mechanical Ability	O Rourke Mechanical Aptitude	Men in Gen l	53
Spatial Relations	Minn Paper Form Board, Rev	Coll Fresh	3
Personality	Calif Self Adjustment	"	95
	Social	"	90
	Total	"	95

VOCATIONAL INTERESTS

Biological Sciences	B	Social Sciences	B−
Physical Sciences	A	Business Detail	C +
Technical		Business Contact	B+
Carpenter	C +	Literary	B
Policeman	B+		
Farmer	B		

Paul's school record showed a high level of achievement his grades all being 85 or better and the bulk of them 90 His achievement in verbal subjects was slightly higher than grades in quantitative subjects but the difference was not great enough to be conclusive He had been given the Otis Quick Scoring Test of Mental Ability Gamma Form, at the beginning of the 11th grade, and had been given an I Q of 113 it was noted on the cumulative record however, that he had ranked 9th in a class of 185 pupils, suggesting that there might have been something wrong with the testing The raw score was 52, a recheck of the I Q shows that this is the equivalent of an I Q of 113 Other notations showed that he was very well thought of by the school staff, both as a student and as an athlete

The problem as Paul saw it was to find a way to finance a higher education, and to decide in which field to major Although attracted to both engineering and to law, he really had no definite idea as to what he wanted to do As engineering specialization begins early, he felt the need to choose or reject it during the last year in high school

Exercise 4

a) Prepare a written analysis of the test results of this case as though for transmittal to his grade adviser Use the sample on page 582 as a model Do this *before reading further and save your report to compare with the appraisal* actually made by the counselor

b) Outline the plans which you think are most suitable for this client, including your approach to counseling the client, in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

James G Revere A Case of Dissatisfaction and Desire to Change Occupations

Mr Revere was a 29 year-old credit clerk, single, a graduate of an academic high school in the small city in which he was working at the time he first came for counseling He was of average height stocky, and getting slightly bald around the temples in a way that made him look older than his age He was personally attractive, open in his manner and fluent of speech with an interesting touch of humor and cynicism He dressed conservatively and well

The client's first full time job (after high school graduation) was a short lived position as proof boy in a publishing house, after which he left home to take a brief course in diesel engineering For the next six months he was unemployed, then he took a temporary position with his present employers He had been with them since, except for a period of military service He took some correspondence work in diesel engines while in uniform The work as credit clerk was satisfactory insofar as pay and stability were concerned, but held no particular challenge, it seemed like a blind alley

The problem as Mr Revere saw it in the first interview, was to "discover what I am best suited to do," so that he might plan a suitable program of night school study and prepare for a more promising occupation He was not certain just what he wanted from this future occupation, but suggested three things which seemed important to him a

substantial income, satisfying work, and opportunity to make use of his mechanical and clerical training

Exercise 5

a) Prepare a written analysis of the test results of this case as though for transmittal to his adviser Use the sample on page 582 as a model Do this before reading further and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client in

FIGURE 23

TEST PROFILE JAMES G REVERE

Test	*Norms*	*Percentile*
Wechsler-Bellevue Adult Intelligence Test		
Total I Q —123	General Population (adults)	95th
Performance I Q —110		
Verbal I Q —131		
Otis Quick Scoring Test of Mental Ability (Gamma A)		
I Q —127	H S College Population	99th
Michigan Vocabulary Profile Test		
Human Relations	Grade 12 students (At Mean for Bus Admin Srs)	98th
Commerce	College Seniors (+1 above Mean for Bus Ad S)	84th
Government	Grade 12 students (At Mean for Bus Admin Srs)	98th
Physical Science	Grade 12 students (+1 above Mean for Bus Ad S)	96th
Biological Science	Grade 12 students (−4 below Mean for Bus Ad S)	50th
Mathematics	Grade 12 students (−3 below Mean for Bus Ad S)	84th
Arts	Grade 12 students (At Mean for Bus Admin Srs)	84th
Sports	Grade 12 students (+3 above Mean for Bus Ad S)	98th
Minnesota Test for Clerical Workers		
Numbers	Male Clerks	16th
Names	Male Clerks	40th
Pennsylvania Bi-Manual Work Sample		
Assembly	Male Industrial Workers	23rd
Disassembly	Male Industrial Workers	31st
Minnesota Paper Form Board (Rev Ed)	H S Graduates who were applicants for a technical course	38th
	Jr-Sr Vocational School students	50th
O'Rourke Mechanical Aptitude Test	Applicants for Mechanical Jobs	93rd
Bennett Mechanical Comprehension Test	Candidates for Technical Courses	87th

FIGURE 23 (Continued)

	Strong Vocational Interest Blank for Men	*Letter Score*	*Kuder Preference Record*	*Percentile*
I	Scientific		Scientific	39th
	A Biological			
	1 Physician	C+		
	2 Dentist	C+		
	3 Psychologist	C		
	4 Artist	B−	Artistic	91st
	B Physical			
	1 Architect	B		
	2 Engineer	B+		
	3 Mathematician	C+		
	4 Chemist	B		
II	Technical Mechanical		Mechanical	75th
	1 Production Manager	B+		
	2 Math Science Teacher	C+		
III	Social Welfare		Social Welfare	3rd
	1 Y Secretary	C		
	2 Personnel Manager	C+		
	3 City School Supt	C		
	4 Social Science Teacher	C		
	5 Minister Rel	C		
	6 YMCA Physical Dir	C		
IV	Business Detail			
	A Clerical		Clerical	81st
	1 Office Worker	B−		
	B Computational		Computational	83rd
	1 Accountant	B−		
	2 Purchasing Agent	A		
V	Business Contact		Persuasive	82nd
	1 Sales Manager	B		
	2 Life Ins Salesman	C		
	3 Real Estate Salesman	B−		
VI	Literary Legal		Literary	70th
	1 Author Journalist	B−		
	2 Advertising Manager	B		
	3 Lawyer	C+		
VII	Miscellaneous			
	1 C P Accountant	B−		
	2 Musician	B	Musical	48th

cluding your approach to counseling the client in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

Ruth Ann Desmond A Case of Dissatisfaction and Ill Defined Objectives

Miss Desmond was a 23 year old high school teacher of business subjects, a tall, very slender, shy young woman with a warm smile which appeared frequently during interviews She had graduated from the state

university and subsequently taught for two years It was during her third year of teaching that she came to the counselor In college Miss Desmond had been most interested in mathematics, and had chosen accounting as a practical application of her interest She had worked during her summer vacations, including those after she graduated, as an electrical unit assembly girl, salesgirl, stenographer, and finally junior accountant Her brief experience in this last field had been of a routine nature, and led her to conclude that she did not want that kind of work Investigation of the work of her associates in the accounting office did not improve the picture, even with promotions in mind Her present job as commercial teacher appealed to her no more than the others the students did not seem really interested in business, and this made teaching an unrewarding activity

The client's other activities and interests included miscellaneous social activities, reading historical novels and photography Her older brother

FIGURE 24

TEST PROFILE RUTH ANN DESMOND

Test	*Norms*	*Percentile*
A C E Psych Exam	College Fresh	79th
Quantitative	" "	77
Linguistic	" "	81
Co-operative Gen l Culture		
Current Social Problems	" "	94
Mathematics	" "	91
Science	" "	87
Social Studies	" "	85
Literature	" "	84
Fine Arts	" "	65
Minnesota Clerical		
Names	Clerical Workers	25
Numbers	" "	21
Purdue Pegboard	One Trial	
Right hand	Factory Applicants	87
Left hand	" "	81
Two hands	" "	40
Assembly	" "	85
MacQuarrie Mechanical Ability		
Total		70
Dotting		99
Tapping		75
Tracing		55
Copying		64
Pursuit		53
Blocks		60
Bennett Mechanical Comprehension		
W W	Waves	70
Minnesota Spatial Relations	Civilian Adults	C—

FIGURE 24 (Continued)

Interest *Strong's Blank*	*Grade*	*Kuder*	*%ile*	*Allport-Vernon*	*%ile*
Author	C	Literary	83		
Librarian	C				
Artist	B	Artistic	60	Aesthetic	30
Physician	C	Scientific	57	Theoretical	30
Dentist	C	Mechanical	47		
Life Insurance Saleswoman	C	Persuasive	88	Economic	50
Social Worker	B+	Social Service	53	Social	60
English Teacher	B+				
Social Science Teacher	B+				
Lawyer	A				
YWCA Sec'y	C			Religious	75
Math Sci Teacher	C	Computational	75		
Nurse	C				
Stenographer	B+			Political	75
Office Worker	B+	Clerical	63		
Housewife	C				
		Musical	49		

Bernreuter Personality Inventory		*Minnesota Personality Scale*	
Emotional Stability	75th %ile	Morale	53rd %ile
Self Sufficiency	55	Social Adjustment	35
Social Dominance	65	Family	20
		Emotional "	55
		Liberalism	80

had a darkroom which made it easy for her to pursue an interest in developing and printing as well as in taking pictures

Miss Desmond was aware of no clear cut vocational interests She had always expected to go into teaching because that was what her mother talked about for her She said she really did not know what she wanted to do, but when asked whether her real interest might be in marriage rather than a career, it seemed clear that she did, for some time at least, want to make her own career

The problem, as Miss Desmond put it, was "to find out for what type of work I am best fitted" Dissatisfied with the only two applications to which she thought she could put her college interest in mathematics, unaware of any special interests and ambitions at the moment, she wanted help in developing a better understanding of her abilities and their vocational uses

Exercise 6

a) Prepare a written analysis of the test results of this case as though for transmittal to her adviser Use the sample on page 582 as a model Do this before reading further, and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client, including your approach to counseling the client, in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

James L Johnson A Moderately Successful Man in Search of New Worlds

James Johnson was a 36-year-old married man, tall, well built, and athletic in appearance, with a dignity beyond his years He had been employed on government projects as a civilian during the war, having been physically disqualified from military service With the closing down of war plants he was soon to be released from this work, and wanted to start making definite plans for the transition back to peacetime employment

Mr Johnson had graduated from an outstanding technical institute with a degree in business administration, at about the time when college graduates were finding that the depression had made radical changes in the employment situation as they had understood it when they chose their major fields He had originally planned to become an engineer, but had swerved from this objective because of the pessimism of an older friend His first job after graduation was in a factory, where he was employed with the understanding that he would be trained for an executive position He was soon made foreman in charge of a department, but after some time the training program was dropped because of the depression and the prospects of advancement grew slight Although he enjoyed the production work, the hours were long and the temperature unhealthy in his department, so he left after a year to take a job with a retail clothing company This also was for executive training, but as he could not accept this company s questionable policies he resigned after several months His next position was with a distributor, a family owned concern in which he was given the responsibility for setting up a new department the operation of which, once it was established, was such a routine matter, with so few outside contacts, that it bored him and left him tired at the end of each day despite its easiness There being little prospect of promotion to jobs normally held by the family and its connections, Mr Johnson left to become placement director of a small but well established college This involved a slight increase in pay, and he enjoyed the variety of contacts, the pleasant working conditions, and the educated

people he generally dealt with When the war came he took a leave of absence in order to accept employment on a government project Here too he had executive responsibility, varied duties, and better pay

Mr Johnson's vocational aspirations, as he saw them, were for work as

FIGURE 25

TEST PROFILE JAMES L JOHNSON

California Test of Mental Maturity, Adv Battery

Total I Q	124
Language I Q	123
Non-Language	118

Wechsler-Bellevue Vocabulary Test

Full Scale I Q equivalent 120

Minnesota Spatial Relations Test Engineering Freshmen 94th %ile

Bennett Mechanical Comprehension

AA " Job Applicants 15th "

Interests	*Strong Grade*	*Kuder %ile*
Biological Science		
Physician	C	
Dentist	C	
Psychologist	C	
Physical Science		10
Engineer	C	
Mathematician	C	
Chemist	C	
Technical		96
Math-Sci Teacher	B	
Production Mgr	A	
Artistic		54
Artist	C	
Architect	C	
Social Service		72
Minister	B−	
Social Science Teacher	C	
City School Supt	B−	
Y Physical Director	B	
Y Secretary	B+	
Personnel Manager	B+	
Literary-Legal		30
Lawyer	C	
Advertiser	C	
Author-Journalist	C	
Business Contact		95
Sales Manager	B	
Life Insurance Salesman	B−	
Real Estate Salesman	B−	
Business Detail		
Accountant	B+	7
Purchasing Agent	A	
Office Worker	A	7
Miscellaneous		
Musician	C	76
CPA	C	

varied and with as congenial a clientele as those he had known as a college staff member and government official, with a staff to handle detail so that he could concentrate on policy, development, and other broader matters, and pay equal to or better than his wartime salary He could have returned to his college position, but discussion of this matter with

the college president had made it clear that there would be no possibility of increasing the staff of the placement office and little in the way of salary increases for him despite the institution's eagerness to have him return The client therefore felt that he should systematically canvass other possibilities, to re-establish himself in the best possible way in returning to normal employment

The problem with which this client wanted help was the appraisal of abilities and interests, and the analysis and evaluation of ways in which they might best be put to use to help him find work of an executive type, with congenial (educated) associates and contacts, and at a good salary (defined as $5000 or more per year) He realized that it might be difficult to find unless he made good use of contacts, but he thought that a position as an administrative assistant might give him needed experience in some one line or industry and put him in a position to advance to executive responsibility He was considering in addition selling tangibles such as cars or oil, especially if he could get an agency He was not interested in insurance and other intangibles

Exercise 7

a) Prepare a written analysis of the test results of this case as though for transmittal to his adviser Use the sample on page 582 as a model Do this before reading further and save your report to compare with the appraisal actually made by the counselor

b) Outline the plans which you think are most suitable for this client, including your approach to counseling the client in the light of your psychometric report Save these for comparison with the counselor's conclusions and with the results of counseling

Counselors Appraisals and the Immediate Results of Counselling

In this section the interpretations of test and other data made by the counselors who worked with these persons will be presented, followed in each case with a statement of the immediate outcomes of counseling that is, of the plans decided upon by the client or of the apparent status of his thinking when he left the counselor Readers who wish to derive maximum value from this chapter should, before reading this section, have made notes on their own diagnoses and prognoses as arrived at while reading the preceding section In some cases, in which the amount of specific detail in the case record permits and the techniques of counseling are interesting, material is included to illustrate points made in Chapter 20

Thomas Stiles Diagnosis and Counseling (case material on p 590 ff)

The Counselor's Appraisal Tom's intellectual level, as shown by his Otis I Q of 101 and confirmed by an A C E score which put him at the 14th percentile point of a typical college freshman class, was about average when compared to that of the general population Occupational intelligence norms from both World Wars indicate that this is the ability level typical of skilled tradesmen and of the most routine clerical workers, observation confirmed by various studies made with the Otis test in industry His mastery of school skills and subjects as shown by his scores on the achievement tests was about that to be expected from one of his mental ability level, and decidedly below that of the college freshmen with whom he was compared, except for a superior score on the mathematics achievement test—his favorite subject This suggested that he might have abilities useful in technical occupations at the skilled level which seemed appropriate to his mental ability His school marks, however were not so encouraging, being only B s in mathematics prior to his junior year, and C's since then The explanation may have lain in his being in the more abstract college preparatory course

On the special aptitude tests Tom appeared to lack speed in recognizing numerical and verbal symbols such as is required of even routine clerical workers Combined with his marginal intellectual ability for office work, this strengthened the basis for questioning the choice of a clerical occupation On the other hand, Tom's scores on the tests of spatial visualization and mechanical aptitude seemed to confirm the implications of the mathematics achievement test His inventoried interests, too, were in the physical science and subprofessional technical fields, the latter field seemed more in keeping with his intellectual level and with his poor achivement scores and fair grades in the natural sciences

Tom's family background, leisure activities, and expressed vocational ambitions were all congruent with the implications of the test results His father was a semiskilled worker, indicating that work at the skilled level might well be accepted by the family as a step upward There were no older siblings who might have established a higher record for him to compete with His leisure activities were nonintellectual, but they did show interest and achievement in mechanical and manual activities as well as familiarity with work at those levels He stated that he wanted to work with engines It was true that, under the influence of a college

preparatory course in the academic high school of a substantial middle-class community, he raised the question of going to college to study engineering, but in most contexts his discussions of work with engines were pitched at the skilled level

The counselor who worked with Tom therefore felt that Tom would be wise to aim at a skilled trade, either by means of a technical school of less than college level, through apprenticeship, or through obtaining employment as a helper in an automotive maintenance shop and taking night school courses

The Counseling of Tom Stiles The counselor began by asking the pupil to bring him up to date on his thinking about his postschool plans Tom did this, indicating no real change in his ideas and mentioning college only incidentally to rule it out as an impractical objective The counselor then reviewed the evidence of the tests and school grades, discussing the intelligence test data in terms of general population percentiles and college freshmen, but focusing mostly on their occupational equivalents Family socio-economic status and the low intellectual level of the boy's leisure interests were of course not mentioned the acceptability of skilled work to Tom and his family was considered something for him to mention, if at all, and as an attitude for the counselor to accept, reflect, and clarify Tom did not mention it, seeming to consider it quite acceptable His leisure activities were mentioned by the counselor as fitting in with the aptitude and interest data, this interpretation was accepted by the client with the statement that "Yes, I always have thought I was best at mechanical things, and I like them best, too"

Ways of utilizing Tom's skilled technical potentialities were taken up next No decisions were reached in this interview, but two nearby technical schools were discussed, and the counselor made sure that Tom had access to their catalogues, one of which was examined in order to review admission requirements, courses, and expenses, and to be sure that the student was oriented to such matters The apprentice training program of a nearby factory, in which aircraft engines were being produced and increasing numbers of young men were being trained, was considered Tom knew about it, and discussion helped him to plan how and when to apply if he decided on it, he saw the possible advantages of having such a specialty if he were drafted Less formal ways of getting experience with automotive engines were looked into, and night schools offering appropriate courses were mentioned Tom was not sure what he would do,

when he left the office, but he felt that he knew a number of suitable alternatives and that he could choose between them in good time

Exercise 8

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question, or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What shortcomings are suggested, in your work or in that of the counselor? Evaluate these in the light of the client s reactions and the immediate outcomes of the counseling as it was done

Marjorie Miller Diagnosis and Counseling (case material on p 592 ff)

The Counselor's Appraisal Marjorie's scholastic aptitude tests indicated that she would probably stand in the top quarter of a typical college freshman class, although they did not justify the principal s characterization of her as "brilliant " Her vocabulary and reading scores suggested that this characterization might be based in part upon unusual ability to put her aptitudes to work, for her reading speed was decidedly superior to her scholastic aptitude and even to her vocabulary level Marjorie was not outstanding on the social studies achievement test, in which subject she had had little preparation, and only moderately superior in the natural sciences which appealed to her, but this latter may have been due to not having included physics in her program Her mathematics achievement was very superior In general, these data were in keeping with the school grades, which we have seen to have been superior, but the trends were reversed, for her mathematics grades were slightly inferior to those in verbal subjects

Marjorie s perceptual speed, when working with clerical symbols, was in the average range for clerical workers, her standing on the numbers test being high average and on the names test low average Her score on the test of ability to visualize spatial relations was only moderately high for college freshmen, and would therefore not be outstanding when compared to scientific workers It seemed high enough, however, to warrant no special consideration if other things were favorable

The personality inventory scores revealed nothing of significance Her interests, as measured by Strong s Blank and the Allport Vernon Study of Values, seemed to be concentrated in the scientific field, with some signs of interest in the social and religious fields They were quite feminine, and although she did not have much in common with office workers who tend as a group to make high housewife scores, her interests did resemble those of housewives It seemed worth noting that her highest scientific interest score was as nurse, which hardly belongs in that group and which is heavily saturated with the interest factor which is most important in housewives She had stated that she thought she might eventually marry, but this thought seemed to play no part in her vocational planning

Marjorie's school and leisure time interests did not do much to weight the balance in the direction of either scientific or social interests Her favorite school subjects included chemistry, languages, and history Her activities encompassed not only photography, but also the school paper, scout leadership, drama, and painting

Marjorie's expressed ambitions were in the direction of natural sciences, in keeping with her measured interests, tested achievement, and some aspects of her school and recreational record The counselor was inclined to give more weight to these factors than to the secondary interest pattern in social welfare work the social welfare and literary activities, and the achievement in verbal subjects in school He concluded that Marjorie would be wise to go to a liberal arts college where she would still have opportunity to explore both the social welfare and the scientific fields in courses and in activities He thought that it would be well for her to select a college which had strong offerings in the natural sciences so that if she did choose this field she would be able to prepare for it as well as her abilities and drive warranted It was the counselor's opinion that Marjorie would probably become a medical laboratory technician, dietician, or high school teacher of science

The Counseling of Marjorie Miller Like that of Tom Stiles, the counseling of Marjorie Miller was done in a situation in which one or at most two interviews were customary, case material being worked up ahead of time and discussed in a factual manner with the student During and especially after the review of the data by the counselor, in terms of their actuarial significance for educational and vocational choices, the pupil had opportunity to react to them and to discuss them The counselor then attempted to help the client understand his reactions, see the

implications of the data, and consider possible lines of action He drew on whatever informational resources were needed and available in order to help with the pupil's orientation In Marjorie's case the review of the data seemed to bring out little that she did not already realize, although the objective and actuarial form in which they were presented impressed her, as might a view of oneself in a mirror for the first time in one's life The possibility of keeping her program broad for the first two years of college and still finishing with a vocationally usable major had not been known to her, when this was mentioned, she suggested that she might then do well to continue to explore both the scientific and the social fields before making a decision

Marjorie then raised the question of which college, as those she had been thinking of, rather vaguely, did not offer genuine liberal arts programs but instead specialized from the freshman year The counselor mentioned several colleges of the type which he thought might be appropriate to her, and asked Marjorie if she had ever thought of any of them Finances appeared to be a problem The counselor had made note of some scholarships for which students might possibly apply, one of them being a very desirable scholarship offered by a first rate college and limited to girls from her part of the state Marjorie wondered whether she could qualify for such a prize The counselor, knowing the standing of some girls who had previously been awarded it said he thought she might and encouraged her to apply She decided to do so, although she could not afford to go to that college without generous financial aid There was some discussion, also, of ways in which campus activities, courses, and summer vacations could be utilized by Marjorie to get a better idea of the direction in which she wanted to turn when she came to the fork in the road

Exercise 9

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question, or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What shortcomings are suggested, in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

Ralph Sheridan Diagnosis and Counseling (case material on p 595 ff)

The Counselor's Appraisal The psychometric data indicated that Ralph was indeed college caliber, having more scholastic aptitude than about 90 out of 100 college freshmen His vocabulary and reading ability were on a par with his promise, perhaps superior to it His mastery of school subjects as measured by the achievement tests was also superior, his greatest strength being in the social studies, with mathematics and natural sciences significantly lower but also better than those of most college freshmen This agreed with his high school grades as to level of accomplishment in general, but revealed differences in mastery which were greater than the slight trends shown in his grades

Ralph's ability to perceive numerical symbols quickly and accurately was low when compared to that of clerical workers, but his perception of verbal symbols was superior In ability to judge shapes and sizes Ralph equaled the typical college freshman but showed no special promise in comparison with freshman engineers In understanding of the uses of the tools and materials of mechanical and related work, he was superior to the typical skilled worker This is not surprising in one who led an outdoor life, his preference for woodcraft activities rather than mechanical may be related to the lack of a high degree of spatial comprehension and might lead one not to emphasize the mechanical aptitude score

The somewhat low adjustment scores fit in with the solitary and small-group leisure time pattern, but are not low enough to give cause for concern

Ralph's interests, as assessed by Strong's group scales and a few supplementary individual keys, most resembled those of men in business contact occupations such as life insurance sales They resembled somewhat those of men in engineering occupations, clerical and accounting work, and the literary-legal fields, but less so They were rather like those of farmers, and may be presumed to have been rather like those of forest service men These interest scores give some support to his expressed desire to study engineering, but, combined with his lack of technical hobbies superior mastery of the social studies, unusual verbal ability, and preference for English, historical fiction, and adventure stories, give even more reason for questioning the choice of engineering The counselor believed that Ralph might graduate from engineering school, despite the fact that many like him drop out or change fields he doubted very much whether Ralph would use engineering training in earning a living His solitary

interests, however, made business contact work seem unlikely to prove satisfying If Ralph was set on engineering, it seemed wise to suggest that he consider industrial engineering, production management, and similar activities rather than the more technical aspects of the work Business administration might be a better major Forestry seemed like another possibility which might give him a combination of the things which interested him

The Counseling of Ralph Sheridan The data were reviewed with Ralph as they had been with Tom and Marjorie, on an information-sharing basis He saw the reasons for questioning his choice of engineering, but felt that he still wanted engineering training He believed that as a civil engineer he might be concerned mostly with the management of production or construction work, and that this would be in line with his semitechnical interests He expressed an interest in exploring nontechnical uses of engineering training as he progressed in his training The counselor felt that he had a good grasp of the situation, and that better insight might develop later The rest of the interview was devoted to places at which Ralph might obtain the desired training and ways of financing it, problems with which we are not here concerned

Exercise 10

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question, or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What shortcomings are suggested, in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

Paul Manuelli Diagnosis and Counseling (case material on p 597 ff)

The Counselor's Appraisal Paul's scholastic aptitude tests confirmed the opinion that the earlier I Q did not truly represent his mental level Compared to college freshmen he seemed very superior, ranking in the top 15 percent His vocabulary and reading speed were even higher The achievement tests showed that he was unusually well prepared in mathematics, superior in social studies, but not well prepared in the natural sciences This seemed surprising, as he had received an 85 in

chemistry the preceding year, and an 85 in the first semester of physics at the time of testing, as his grades in the linguistic subjects were generally slightly superior to those in the quantitative, this may actually reflect true differences in his special abilities

In ability to distinguish numerical symbols with speed and accuracy Paul ranked low when compared to clerical workers, but his facility with verbal symbols was average for such workers His understanding of the nature and uses of mechanical woodworking, and related tools and processes was average when compared with that of skilled workers, but his ability to visualize and mentally manipulate objects in space was quite inferior when compared with that of college freshmen Despite the high achievement in mathematics, this poor showing in spatial visualization and relatively low standing in natural sciences, combined with high verbal ability and superior social studies achievement, appeared to lend some support to this student s second expressed interest law His very superior measured adjustment agreed with the opinions of the school staff

Paul s inventoried interests were most like those of physical scientists, including engineers, and also resembled those of men in business contact work such as life insurance sales He showed some interest also in the biological science and literary legal occupations, but not as much as might have been expected of a boy who enjoyed public speaking and debating had little in the way of technical hobbies and was considering law as seriously as engineering

The counselor was inclined to give more weight to the factors pointing in the direction of engineering than to those contradicting it Mathematical ability and interest supported the choice, while poor achievement in science and poor spatial visualization opposed it It seemed possible that the spatial relations score was for some reason not representative, and the poor science preparation might have been more a matter of teachers than of pupil It hardly seemed justifiable to question seriously the choice if Paul made it after a review of the data

The Counseling of Paul Manuelli In view of the above, the counselor let Paul talk some about his vocational objectives These seemed more than ever to involve engineering training, but Paul wanted to know how he compared with engineering freshmen His profile was therefore reviewed Paul reacted particularly to the relatively high mathematics standing, to the low Paper Form Board score, and to the greater degree of interest in physical science than in legal occupations He was not

inclined to be discouraged by the low spatial score, and might perhaps have given it no more thought, but the counselor suggested that it might be taken as a warning signal, and that if mechanical drawing or other such activities ever gave him trouble he might want to look into it further, it was mentioned also that his standing might be checked by a retest College choices were then discussed, the client raising the question after indicating that he thought he would go ahead with engineering Paul felt that his financial status might make choice of a co-operative training program wiser than a four-year engineering school The nearby engineering schools operating on the co-operative system were therefore considered with the aid of catalogues, and one was most thoroughly discussed as being accessible, inexpensive, and of good standing

Exercise 11

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What shortcomings are suggested, in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

James G Revere Diagnosis and Counseling (case material on p 599 ff)

The Counselor's Appraisal Mr Revere s intellectual ability, as shown by an individual and a group test of mental ability was decidedly superior, both tests placing him above the 95th percentile A test of specialized vocabulary revealed that he had unusual verbal ability in a variety of fields Although none of the norms for this test were strictly appropriate to this client, who was much older and more experienced than the high school seniors and less trained in business than the college business seniors to whom he was compared, the indications were that he was well informed in all fields except the biological sciences

His speed in handling clerical symbols was poor for numbers and average for names, when compared with clerical workers His manual dexterity was fair, as was his ability to judge shapes and sizes and mentally to manipulate them His knowledge of mechanical tools and processes, and his ability to comprehend mechanical principles and opera-

tions were very superior when compared with those of persons with some experience and interest in those fields

The client's interests, according to Strong's Blank, were most similar to those of men engaged in scientific, subprofessional technical, and business detail occupations It was notable that none of these patterns were clear cut, each involving some B— or C+ scores, and only one, the business detail group, including an A (purchasing agent) At the same time, there were scattered B's in several other fields, including business contact and literary legal occupations The Kuder pointed these findings up by yielding a low scientific interest score, with a high mechanical, even higher scores were made, however, in computational, clerical, and persuasive interests The high artistic and substantial literary scores on the Kuder were discounted as lay interests as they were not appreciably reflected in the Strong or in the client s leisure activities

The test results generally appeared to fall into no more clear-cut a pattern than did the interest inventories but some study of them suggested a few conclusions of significance The combination of interest in mechanical or subprofessional technical work with interest in persuasive activities and mechanical aptitude and fair spatial visualization suggested that sales and service in the mechanical field might provide a suitable outlet for the client Contraindicating this, however, seemed to be the client's intellectual level and his desire for status

These same low level or attenuated technical interests and abilities being also characteristic of production managers and industrial engineers, who are characterized by a level of mental ability more nearly resembling that of the client, it seemed to this counselor that work of this type might provide Mr Revere with activities of a more satisfying type accompanied by status more in keeping with his desires

While the testing was going on, however, the counselor had a number of interviews, extending over a period of three months, with the client These interviews were focused at times on types of work and ways of getting into them, but at other times on the client s feelings of insecurity These were brought out only incidentally in the second or third contact, although the counselor had suspected their existence in the first interview as counseling progressed they came to the surface more often, and were recognized as a fundamental part of the client s adjustment problem

The counselor's diagnostic formulation of the case was as follows Mr Revere was faced with a very real problem which is not uncommon

among clients in the late twenties and early thirties It was that of an intelligent, maturing man who had not developed or achieved any clear-cut occupational goal and was becoming concerned about it He evidently realized that he needed to take steps of some sort in order to derive greater satisfaction from his work Having held only one real job, and feeling that he knew little about occupational opportunities, he sought the help of the counselor His feelings of insecurity (unrecognized at first), together with his overconfidence in the prescriptive power of tests, caused him to expect the counselor and tests to solve his problems for him He was therefore reluctant to make his own decisions and take the responsibility for his own actions

The Counseling of James Revere The counselor believed that there were two important objectives in his work with Mr Revere One was to help him to clarify his objectives, values, and interests by discussing these at length in a permissive and insight-producing situation The other was to help him accept, understand, and overcome his feelings of insecurity, by letting him talk about them and discover ways of handling them As the client had come to the counselor with vocational aspects of his adjustment as his reason, and as the objective data indicated that he had some reason for feeling misplaced, it was felt that the best way to get at values, life goals and feelings of insecurity, was through a discussion of the client's vocational problems From the beginning the client felt strongly that tests would help him, despite an attempt to play them down during the first two interviews, it was therefore decided that testing might be a help in keeping some kind of rapport, and that the counselor's skill would be most effectively used in helping the client to see the importance of other factors after rather than before testing

The first four interviews were therefore devoted to discussions of vocational goals and opportunities and of the client's feelings of insecurity, and to test administration and interpretation, the testing being done by the counselor as part of an interview By the fifth interview the client began to show some interest in following up an old objective, sales and service work with business machines He felt that this would pay well and use both aspects of his previous training He felt, however, that his feelings of insecurity in relations with other people would be too great a handicap He showed considerable dependence on test scores, on the counselor, and on two friends whom he considered successful and well informed The sixth and seventh interviews were devoted to discussion of the somewhat low clerical perception, manual, and spatial scores,

which rather disturbed the client and to exploration of possibilities in mechanical and advertising fields This exploration was almost entirely a matter of discussion in which the counselor quizzed the client in order to get him to tap his own sources of information or supplied information himself, and in which he reflected the client's feelings in such a way as to help clarify his attitudes toward the opportunities under discussion

The turning point at which Mr Revere began to assume some responsibility for solving his problem himself, a point long recognized as crucial in psychotherapy but generally unrecognized in vocational counseling, came in the eighth interview By this time the client had evidently reached the point at which he perceived that tests had helped him as much as they could They had shown him certain unsuspected weaknesses and some half suspected strengths but they had not solved any problems He knew more about himself but the decisions concerning himself still had to be made The counselor had consistently refused to make them for him not by saying in so many words that he must live his own life, but by discussing problems and clarifying feeling in such a way as to leave the responsibility for the last word always in the client's hands

During the ninth interview Mr Revere showed some discomfort and wandered considerably not liking the fact that mechanical aptitude, which might mean beginning again at the bottom of an occupational ladder, seemed his principal asset But he assumed the responsibility for concluding that if he was to get beyond the point which he had reached in clerical work he would probably have to change fields and he made the decision that it should be something mechanical He felt that sales and service would be the logical combination as he had some contacts which might help him get started and it would not mean trying to get the college engineering training that he lacked In the next interview he brought to the surface the feeling that his emotional insecurity was the big stumbling block in sales work This fact had been mentioned before but this time he began to examine its foundations He vacillated between the opinion that he could sell if he tried and that he was so afraid of people that he would not be aggressive enough By the 11th interview his defenses were down for he then clearly realized that the only real obstacle to doing what he felt he should do was his own personality problem The counselor was accepting recognized the nature of the dilemma and discussed it with the client, but did nothing directly to resolve the issue

By the 12th interview the client seemed to have worked things through himself, helped by the impetus gained in discussion with the counselor He discussed his proposed plans with the counselor They included some refresher work on business machines, to strengthen his case in applying for a sales job Then he planned to talk with some of his contacts, to see what further orientation they could give him, especially in the way of places and persons to which he might apply Then he planned to take time off from his job in order to carry out a thorough going job seeking campaign He would look for a job in which he might round out his training in the use and maintenance of business machines before undertaking to sell in the field He still manifested some doubt as to his sales ability, and this was gone into again It came out that his fears had to do with initial contacts, he saw that in work such as this he would in due course reach the point at which there was not too much new-contact work And he was sure enough of his relations with people he knew to be confident that he would do well with regular customers, and even with new customers who were not seen without previous cultivation This seemed correct to the counselor, as the case history material showed him to be a likable and liked young man

The counselor had thought that this interview might close with a decision on the part of the client to undertake psychotherapy, as a necessary preliminary to vocational adjustment The counselor was therefore prepared to handle the transition to another counselor It seemed however, that the client had taken things into his own hands, and assumed responsibility for his own acts Apparently he was not sufficiently uncomfortable about his fears of meeting people to want to explore that matter any further than he already had, with the counselor's help, in the interviews on vocational adjustment The case was therefore closed by mutual consent and on the initiative of the client, after twelve contacts involving interviewing and testing

Exercise 12

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What

shortcomings are suggested in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

Ruth Ann Desmond Diagnosis and Counseling (case material on p 601 ff)

The Counselor's Appraisal Miss Desmond exceeded 79 out of 100 college freshmen on the A C E Psychological Examination, with almost equal scores on the linguistic and quantitative subtests As she was several years above college freshman age and scores are influenced by those years her actual ability was probably not as superior as this test suggests, but in any case she was clearly of superior mental ability On the general culture test the client's highest scores were in current social problems and in mathematics, both of these being in the top decile However, her knowledge of science, social studies, and literature were each almost as high, and her familiarity with the fine arts was also better than average All that could be concluded from this test was that the client was a well informed young woman, it helped little if at all with differential diagnosis

Miss Desmond's ability to perceive numerical and verbal symbols was only mediocre when compared with employed clerical workers, which may have had something to do with her dislike of the junior accountant s work Her manual dexterity in single handed operations was superior to that of women industrial employees in two handed operations she was about average Her speed in dotting and tapping operations of a manual dexterity type which are important in mechanical and office jobs such as machine bookkeeping work, was very superior She was superior also in her ability to comprehend mechanical principles and apply them to operations Her ability to judge shapes and sizes, however, was low average when compared to the general population

The client s interests, as measured by both the Kuder and the Strong, were very much like those of women engaged in social work, teaching social studies and English, and office work These were supported by the Allport Vernon Study of Values, which also brought out considerable interest in status and prestige The Bernreuter and Minnesota personality scales agreed in describing Miss Desmond as emotionally stable, the former adding self sufficiency and social dominance, and the latter pointing up poor family relations, observation of the client led the counselor to consider the Bernreuter scores indicative of compensatory attitudes

and behavior, and the Minnesota scores more truly suggestive of underlying attitudes

The counselor felt that Miss Desmond's assets consisted of her superior mental ability, drive, superior general information, manual dexterity, and mechanical comprehension Her clerical perception seemed good enough to be usable as a means to something else, even though it hardly seemed likely to make for success in clerical work as such The problem seemed to be one of finding ways in which these abilities could be used which would be congruent with her social and office work interests Two possibilities occurred to the counselor 1) statistical machine work, in which the client's mechanical comprehension, manual dexterity, and mathematical ability could be combined with office work interests, perhaps of a supervisory nature which would provide outlets for her interest in dealing with people, 2) secretarial work, for which also she had the necessary training, in which her mediocre clerical aptitude would be more than compensated for by her intelligence interest in human relations, and, perhaps, ability to assume responsibility as an administrative assistant or junior executive

The Counseling of Ruth Ann Desmond In discussions with this client the focus was at first on the reasons for dissatisfaction in her previous employment Testing was done in a supplementary way concurrently with the interviews, by another person, it being made clear that the counselor thought of them merely as another way of getting some information which might be useful When the test results were available the counselor explained their psychological and occupational significance, leaving time for Miss Desmond to express her attitudes and feelings as he did so

The client suggested that perhaps getting a position as a secretary in an office in which she might have a variety of responsibilities, including contact with the public or supervision of others, and rise to more executive types of responsibility might be one outlet for her The counselor reflected the feeling that this might be a good type of opportunity, and it was discussed further The counselor then asked if Miss Desmond had ever thought of work with statistical machines, and found that she knew little about the opportunities in that field These were therefore outlined by the counselor The client left with the intention of exploring both fields

At the next interview she reported that she had been offered several stenographic jobs, one of them being in a law concern with a large and

varied practice She believed it offered some possibilities, and said she might take it Her thoughts seemed to be almost entirely on this matter, and the interview produced little else The next report, by telephone, indicated that she had taken the law concern job and was beginning work

Although this case came as one of vocational counseling, with an immediate problem of vocational choice which caused both counselor and client to focus on pertinent attitudes, aptitudes, and interests, the counselor was not quite satisfied with his work It was true that the diagnostic picture was not clear and that despite this a coherent picture of abilities and interests had been constructed which made psychological and occupational sense, that the immediate outcome had been the making and launching of a vocational plan in keeping with this picture, and that all of this suggested effective work along appropriate lines Despite this the counselor felt uncomfortable about the case He wondered whether it might not be that Miss Desmond really needed help with a problem of personal adjustment, but had been unable to ask for such help or even to take it when the counselor asked rather directly about her ideas concerning marriage It seemed possible that she might even have taken the law office job as a means of putting an end to the counseling relationship in which she might soon have approached her personal problem If so the counselor wondered, might he so have handled things earlier in the relationship as to have avoided such a break? To have probed and rushed right into the problem would hardly have improved things But a focus on attitudes and values rather than on vocational interests and aptitudes might have led more rapidly to the development of a relationship which could have withstood the strain of the uncovering of emotional problems Only a follow up and the subsequent history of the case could tell and even it might be as unconclusive as many of its other features were

Exercise 13

a) Compare your interpretation of the test results with that of the counselor, and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question, or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the counselor Compare your proposed approach with that used by the counselor What

shortcomings are suggested in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

James L Johnson Diagnosis and Counseling (case material on p 604 ff)

The Counselor's Diagnosis Mr Johnson was given two tests of mental ability, the Vocabulary Test of the Wechsler Bellevue and the California Test of Mental Maturity On the former his intelligence quotient was 120 on the latter his total I Q was 122, with language and nonlanguage I Q s of 123 and 118 respectively The evidence therefore agreed in showing him to be a man of superior mental ability, quite capable of performing successfully in professional or executive work

In ability to visualize the relations of objects of different shapes and sizes Mr Johnson exceeded even the majority of engineering students, standing at the 91th percentile on the Minnesota Spatial Relations Test In ability to understand the operation of mechanical contrivances and to apply mechancial principles to practical situations he did not, however, compare well with graduate engineers, his score being at the 15th percentile for this group Although scores on this test are somewhat affected by experience its effect is not very great, and in any case the client had had experience which had given him opportunity to increase his familiarity with mechanical matters and to apply his spatial visualization ability to mechanical problems

Interests were measured by the Strong and Kuder inventories They agreed in revealing a high degree of interest in subprofessional technical occupations such as production manager, occupations which provide outlets for mechanical interests but do not require a high level of mechanical ability or of interest in scientific matters The Strong Blank showed some resemblance between Mr Johnson's interests and those of successful salesmen and sales managers, but not as much interest in sales work as was suggested by the very high persuasive score on the Kuder Record This seemed to be related to the client's statement that he was interested in promotional activities, but did not like actual selling Strong's Blank showed considerable similarity of interest with those of men employed in business detail work, including accountants and purchasing agents, but the Kuder yielded very low scores on the clerical and computational scales Apparently the client had interests like those of office workers but, as he himself stated, did not enjoy clerical routine

once it was established Both inventories revealed some interest in social welfare work, but this seemed secondary to business and production interests

An attempt at personality appraisal was made by means of the Bernreuter Personality Inventory This depicted the client as quite unstable emotionally, dependent, introverted, moderately dominant in face-to-face situations, quite self-conscious, and somewhat solitary These results seemed to agree with interview material which suggested an underlying neurotic tendency in Mr Johnson This material consisted of his description of himself as a worrier, and of a possible interpretation of his vocational dissatisfactions as due to personality maladjustment rather than to vocational misplacement These maladaptive tendencies seemed, however, to be well under control, as evidenced by Mr Johnson's success in each of his jobs, his employers' desires to have him stay with them, and the fact that each change of employment so far had been for a definitely superior position Although it seemed to the counselor that the client might be paying too high a price emotionally for his success, the fact that he did not take advantage of the rather permissive counseling relationship to work on personality problems led the counselor not to press him to open up that area

In summary, it seemed that Mr Johnson was a man of superior general mental ability capable of achieving, as he actually had, at the professional and executive level His low mechanical comprehension and scientific interests indicated that he had perhaps done well to avoid engineering occupations, although he did have the spatial aptitude and lower technical interests which might make industrial work appeal to him This probably explained the satisfaction which he found in the factory job which he held after graduation, despite the poor working conditions and lack of advancement which caused him to leave it The combination of business, technical, and welfare interests shown by the inventories, combined with his own stated preferences, indicated that he should find satisfaction in office work of a supervisory nature, in which he did no detail work but was responsible primarily for planning and for outside contacts It was felt that he might have difficulty adjusting to the emotional demands of some jobs, but his success in his previous positions and in moving to progressively better jobs led to the conclusion that under favorable conditions he would be able to make the necessary adjustments Psychotherapeutic help might enable him to get more satisfaction from

his work and from other aspects of life by relieving him of the load of anxiety which it was suspected that he carried, but it might be more appropriate for him to seek such help after he had made the change back to a peacetime job than during the transition period

The Counseling of James L Johnson In this case the counseling procedure consisted of the initial interview for the collection of case material and the determination of the problem to be worked on, followed by the administration of the tests the results of which have just been summarized Then followed an interview for the discussion of the implications of the test results and of the meaning of the client's experiences to date These were followed by three more widely spaced interviews, in which job seeking plans were thought through, related activities were reported and evaluated, and the suitability of openings discovered was discussed

In the first interview following testing Mr Johnson's test data were interpreted as favoring employment in fields such as production management, personnel work, buying, and general administrative work such as he was contemplating It was suggested that sales work did not seem indicated, and that his former engineering interest might have proved to be an unwise choice had it been followed through After this rather directive interpretation by the counselor the discussion shifted to a review of the client's experiences in the light of the test results, and of the test results in the light of the client's experience In this process the counselor was relatively nondirective, reflecting the feelings and attitudes expressed by the client, and occasionally asking a question designed to assist the client in his thinking When, for example, contemplation of his low clerical score on the Kuder and the moderately high clerical score on the Strong caused the client to remark So I don't like clerical work but I do have interests somewhat like those of office workers' the counselor asked, "What meaning does that have for you?" This led the client to state that he supposed he would find working with that kind of people congenial, but that he would want to have duties other than responsibility for clerical detail The question, "What kinds of jobs might offer you that combination?" led to an exploration of supervisory and public contact jobs in business Further discussion led the client to conclude that, everything considered, the position of administrative assistant would probably offer him the best chance to do congenial work and to learn enough about some type of enterprise to enable him

to assume executive responsibilities Other supervisory and contact jobs did not seem equally open because of lack of specific experience other than clerical and educational

In the next interview the means through which the client might locate suitable openings were explored He revealed a good orientation to job-seeking methods, so the discussion was primarily an opportunity for him to use the counselor as a sounding board for his own analysis of each lead, of the best way in which to use it, and of the suitability of the kinds of jobs which it might yield

Subsequent interviews were devoted to discussion of the openings which the client located during his job hunting One of these was in personnel work with an oil company, others were accountant with supervision of an accounting department for an important foundation, industrial relations work with a rubber manufacturer, industrial engineering in an electrical equipment factory, and administrative assistant to the head of a large business enterprise The two personnel positions had considerable appeal, but both involved certain limiting conditions which made the client hesitate, one geographic and the other the narrowness of the job because of the specialization in the office The accounting job was with an organization which would have provided very pleasant working conditions and good pay, but the client knew that he would have to relearn a great deal about that type of work and that the work itself would not appeal to him The industrial engineering job would have involved beginning rather low in the scale and working up, and at his age and with his experience the client did not feel he should make such adjustments The position of administrative assistant had the most appeal for it not only paid well but was described as one which had for some previous incumbents, led to higher level executive positions in this and in other companies Mr Johnson was quite enthusiastic about this possibility and the counselor felt that it was compatible with his interests and abilities The client stated that if the offer materialized he would accept it

Exercise 14

a) Compare your interpretation of the test results with that of the counselor and note the ways in which they differ Study these differences in order to locate possible inadequacies in your conception of the significance of the tests or scores in question or ways in which your insights may be more adequate than those of the counselor

b) Compare your tentative plans with those considered suitable by the coun

selor Compare your proposed approach with that used by the counselor What shortcomings are suggested, in your work or in that of the counselor? Evaluate these in the light of the client's reactions and the immediate outcomes of the counseling as it was done

CHAPTER XXIV

ILLUSTRATIVE CASES FOLLOW-UP AND EVALUATION

The Validity of Vocational Appraisals in the Light of Subsequent Work Histories

EACH of the seven cases discussed in the preceding chapter was followed up some time after counseling in order to find out in what type of work he was engaged, how well he liked it, what aspects of it he disliked, and how well the ultimate outcomes of counseling agreed with the appraisals made by the counselors The time that elapsed between the closing of the case and the follow up varied greatly In one case it was only three months, as the case was handled not long before this chapter was written In one it was 15 months In several it was two years In some it was six years, and in some it was even longer The cases were, with one exception selected partly because enough time had elapsed since counseling to make follow up meaningful

In one case the follow up was through personal contacts which, by a happy combination of circumstances, were renewed from time to time over a period of several years In certain others it was made through correspondence supplemented by the personal contacts of others living in the same communities And in still others the follow up consisted solely of a brief exchange of letters Such methods leave a good deal to be desired, as they are not likely to yield emotionally toned material and provide insufficient opportunity for the exploration of important issues Their results are, however given for the insights which they do occasionally give into the adequacy of the understandings derived from the tests and related diagnostic procedures Inadequate though they may be, the obtaining of even these follow up data represents an advance over much that is done in the way of test evaluation It is in the intensive follow up and evaluation of clients adjustment that the greatest advances still remain to be made

The follow up data for each of our seven cases are presented in the

paragraphs which follow, accompanied by comments on the adequacy of the testing and of the appraisal in the light of these data

The Early Career of Thomas Stiles (see also p 590 and p 607)

Subsequent History Tom was followed up by means of a personal letter some eight years after he was tested and counseled His letter was brief and factual, giving an outline of his experiences during the intervening years but not going into detail concerning his attitudes toward these experiences After graduation from high school he took a summer job, comparable to those he had held in previous years He was then admitted to apprenticeship training in a large metal products manufacturing concern, remaining for six months before illness caused him to resign and return home Several months later he accepted employment as a tool grinder with a company within commuting distance of his home, where he worked for a period of two years He was successful at this work, but felt the need for more training of the type that had been interrupted by his illness He therefore gave up this job and entered one of the subprofessional technical schools which he had discussed with the counselor three and one half years previously, taking a two year course in steam and diesel engineering He graduated after the normal two years, having enjoyed the training and worked during the one summer in another metal products factory After completing his training he was placed, by the school placement service, in a job with a manufacturer of railway locomotives Nine months had elapsed on this job at the time of writing and Tom felt that he had successfully begun a career of the type which appealed to him most, with a concern which would offer him security and advancement

Validity of the Appraisal The plans carried out by Tom seem to have corresponded rather closely with the appraisal of the counselor insofar as type of activity is concerned, although there was some floundering at the start and the ultimate achievement level was the highest of those which had been deemed likely It will be remembered that the counselor had thought of apprenticeship or on the-job training as equally appropriate in Tom's case as formal training in a technical institute Although the interruption in the apprentice training which he began after completing high school seems to have been due to factors not related to his interests or abilities, the fact remains that it was interrupted, that a period of work followed which served to finance schooling and to confirm his desire for it, and that in the end he graduated from a technical

institute and obtained employment at the skilled level The only way in which Tom's history differed from that discussed in the counseling process was in the selection of steam and diesel rather than gasoline engines, but this difference is, from the standpoint of aptitudes and basic interests, quite insignificant

One might conclude from this one case that test results and the counselor's diagnoses tend to correspond rather closely, were it not for the fact that while a case may illustrate it cannot prove The case of Marjorie Miller, which follows, serves to bring out the complexity of people and of occupations, and to underline the fact that vocational adjustment is often a process of unfolding rather than of predicting

Exercise 15

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

The Early Career of Marjorie Miller (see also p 592 and p 609)

Subsequent History Marjorie carried out the decision reached with the aid of the counselor and applied for the special scholarship at the high ranking college Drawing on the diagnostic data made available by the counselor, the principal gave her an extremely favorable and yet objective recommendation She was awarded the scholarship, which provided all she needed to supplement her family's financial backing, for her four years in college At the end of her freshman year the counselor had a letter from Marjorie, expressing her appreciation of the educational experience which he had helped her to obtain, and describing some of her reactions to her first year Apparently her horizons had been so broadened by the experience that she felt considerable gratitude to the counselor for having made her aware of the advantages of the type of college she was attending and for having found a way to make it financially possible The next contact came at the end of Marjorie's college career when the counselor received an announcement of the graduation ceremonies in which Marjorie was to participate The third follow up

was made through one of the college personnel officers, a year after Marjorie had graduated and six years after the counseling took place

Marjorie's freshman program in college included chemistry, economics, English, and German Her grades for the year were C, C, B and B, respectively Her college personnel record showed that her goal, at the beginning of this year, was "nutritionist chemist, or social worker" *Late in her freshman year she discussed her choice of major field with* a counselor, who was impressed by her intelligence, viewpoint, and enthusiasm She talked also with the heads of the science departments in which she was most interested

During her first summer vacation Marjorie worked as a sales clerk in a department store, and acted as head of the department in which she worked Her employer reported that "Marjorie has better than average intelligence, good initiative, and excellent character While at work in my store she handled selling duties very well although she had had no previous training in this field She is ambitious and would succeed in any work which she undertook

In her second year in college Marjorie apparently shifted from her former scientific inclinations and majored in child study She took four courses in this subject, continued economics and German, and added physiology and psychology Her marks for the year were all B+ or B She continued along these lines during her third and fourth years, concentrating more and more on psychology and child study Her marks improved steadily, and she graduated 85th in a class of about 250 students

Marjorie's extracurricular activities consisted of working on the college newspaper as a reporter during her first year and as assistant managing editor of a new and rival publication during her sophomore year She was an officer, and ultimately president, of a campus religious organization She served as co-editor of her class yearbook in her senior year

In her last summer vacation Marjorie took a position as a playground instructor in one of the large cities, receiving ratings of "excellent" in industry, ability, attitude and attendance She also did field work with children in a local settlement house as part of her academic work during the year Her supervisor's report read "She showed a fine understanding of the needs of individual children, was responsible for completing tasks assigned to her, and showed initiative in many situations where students frequently wait for direction She has a friendly personality and adjusts easily to new situations" Another supervisor spoke of her "good sense of

orientation, quick grasp of problems What is more, she showed good intellectual and emotional insight into the life of children"

When Marjorie registered with the placement office during her senior year she stated that she wanted to teach in a public or private nursery school In a later contact she expressed the same interest, but hesitated about applying for specific openings because she was thinking of marrying soon and really wanted a job more than a career A little later she expressed an interest in an opening as a field secretary for one of the scouting organizations, had an interview with a representative of the national office and was employed in a branch near her home town

A final follow up revealed that Marjorie was doing well in her work, found it very satisfying, and had been promoted to a more responsible position in the same organization Although she still had marriage in mind it had receded into the background at least temporarily, and she looked forward to continuing in the same work for the foreseeable future

Validity of the Appraisal Marjorie's grades in college were in line with the counselor's expectations, when he disagreed with the high school principal's characterization of the girl as brilliant She did prove to be, as he anticipated a good student in her chosen field graduating at the bottom of the upper third of her class It is interesting to note, however, that her work in science and freshman (but not sophomore) economics was only at the C level Her achievement in the more verbal subjects was better than that in the more quantitative as suggested by the analysis of her school grades This trend was not however clearly foreshadowed by the test scores

Of major interest is the predictive value of the interest inventories These, it will be remembered showed dominant interest in the scientific fields with some signs of interest in social welfare and religion Her school and leisure time activities did not do much to decide the issue one way or the other, as they included scientific and social interests The subsequent history showed that, contrary to the counselor's expectation, the secondary social welfare interest pattern became dominant as time went by It has been seen that Marjorie carried out the program of exploration in both scientific and social areas which the counselor had recommended for her freshman year, and that, whether because of interest aptitude or some combination of the two, she then focused entirely on the social welfare field

The counselor's private opinion, then, which he did not let influence his counseling was mistaken He had thought that exploration would

confirm Marjorie in her adolescent choice of a scientific occupation, whereas in fact she decided to prepare for and actually entered a social welfare occupation Stated as flatly as this, the outcome might lead one to conclude that the test results had actually been misleading in this case But reconsideration of the data will reveal that the basis of Marjorie's subsequent actions can be seen in the high school counseling record Subordinate though the trends seemed to the counselor at the time, there were indications of social welfare interests Isolated from the rest of the pattern these indices are rather impressive

Unusual reading speed
Verbal grades superior to quantitative
Secondary social welfare interest pattern
Feminine (i e social and literary) interests
Active on school paper
Dramatic club member
Scout leader

The foundations for the choice of a social welfare or literary occupation were clearly there It was only the more dominant interest in science supported by superior achievement in the sciences and equally important scientific avocations which led the counselor to believe that success and satisfaction were most likely to lie in the applied sciences

Perhaps the principal conclusion to be drawn from this case, however, is that even in the case of some well motivated, clear thinking, able high-school seniors interests and abilities are still in the process of developing or, at least, of coming to the surface of consciousness When more than one pattern of abilities and interests is noted it is therefore wise for the student to plan a program of study, work and leisure which provides for further exploration of the two or three dominant patterns The diagnostic process may serve to reveal areas in which exploration can best be concentrated, and counseling may have as its function the planning of appropriate types of exploratory activities Actual decision making may not come for some time, and then it will turn out to be a step-by-step process rather than an event

Exercise 16

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your

mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

The Early Career of Ralph Sheridan (see also p 595 and p 612)

Subsequent History In response to a follow up letter written six years after counseling Ralph wrote in part

At the time I was planning to enter Renssalaer Polytechnic Institute The tests showed that while I might make a fairly decent showing at engineering, I would probably do better at other things I think they were 100 percent right

I entered Renssalaer but was obliged to withdraw during the third semester by the death of my brother My marks were such that I could have got by but not much more

I then went to work in an abrasive factory at a semiskilled job which I liked rather well but found monotonous Then I entered the Seebees and enjoyed the construction work we did at various Naval installations

Upon my discharge, I went back to the factory as a foreman then rose to assistant to the superintendent Since then there has been a decline in the volume of production, and consequently a reduction in the number of employees I have worked at various temporary jobs in the same plant since then, none of them significant just to keep on working until the beginning of the Fall term

I plan to enter the Syracuse University School of Business this Fall which you suggested (very wisely I now believe) as a suitable plan six years ago I think that your analysis of my abilities and interests was quite accurate and can say that, even though I did not act upon it then it has helped me to understand my subsequent experiences and to make plans based upon my assets as they have been demonstrated in my work

Validity of the Appraisal It is interesting to note that the client has emphasized in retrospect the strength of the suggestion of business administration training made by the counselor and the definiteness of the verdict of the tests ("they were 100 percent right") Such oversimplification on the part of counselees is common, and should serve as something of a deterrent to counselors who tend to be overdirective The data themselves are generally quite sufficiently directive if not, more direction in the form of advice from the counselor may well be harmful

The validity of the appraisal is demonstrated by several facts in Ralph's experience First, there is the mediocre record in engineering school, confirming the diagnosis of weakness in that area Secondly there is the success in the administrative side of production work, which led to promotion to foreman and assistant to the superintendent And

finally there is the client's conclusion, from these and his other experiences, that business administration training would be most in line with his interests and abilities and would best equip him for work of the type he wanted

The case is interesting also, in that it illustrates how a counselee may reject the implications of tests and counseling, proceed to try out his plans, and ultimately revise them to make them conform with the original counseling In Ralph's case, as in many others, the counselee seems no worse off for having found things out 'the hard way", in fact, he may have learned some useful lessons as a result, and may profit more from his subsequent education that he otherwise would have But testing and counseling seem to have been valuable in forearming him, in making it easy for him to learn from experience and to revise his plans as needed Testing and counseling, then, converted floundering into exploration

Exercise 17

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

The Early Career of Paul Manuelli (see also p 597 and p 619)

Subsequent History Paul like Ralph, was heard from by mail six years after he graduated from high school He wrote as follows

I graduated from high school with honors, a medal for excellence in United States History and a scholarship at Carnegie Tech

I spent my freshman year at Tech studying mechanical engineering I was permitted to omit Freshman English taking literature in its place I played on the Varsity football squad I made a C+ average which was all right as my scholarship was based more on athletics than on academic achievement

After we got into the war I joined the Navy V 12 program and was transferred to Stevens Institute, where I graduated with a B S in Mechanical Engineering in 1945 and grades averaging from 75 to 80 I made the Dean's List once in my junior year, played on the football team belonged to the senior honorary society, was class orator, was listed in the American student Who's Who, was company commander, was on various student committees, and took part in theatrical productions rather regularly

Transferred to Columbia Midshipmen s School, I was commissioned an Ensign in the Naval Reserve graduating 259th in a class of more than 1000 men I served as company commander here also

I was assigned to duty on a cruiser as a junior officer in firerooms, machine shop, and main engines Studying under the chief engineer, after seven months aboard I qualified as engineering watch officer and stood watches, in complete charge of the engineering facilities

My discharge came late in 1946 after which I joined the Atlas Corporation as a student engineer I completed a year s study in which I spent several months in each of their main divisions, learning all the operations from design to sales and service After completing this course, I requested assignment to production engineering I could have asked for development engineering, but I felt that I would be better qualified for development work if I became thoroughly familiar with the problems of improving existing designs making them easier to manufacture, etc before trying development work I would like ultimately to be a senior engineer with a department of my own but of course that is a long term objective Before that happened I think I might be tempted to shift to factory administration, as I enjoy working with people and handling long range problems

I enjoyed my schooling and even though I never made top grades I never had any worries about passing I got as much out of college as most students, and never disliked any courses I suppose I cared least for drafting as drawing prints is an anti climax after actually solving a problem The Navy was all right too My work with Atlas has gone smoothly and I have never felt unprepared or unable to handle the work that has come my way

I have had two substantial raises since coming to Atlas and consider my rate of progress satisfactory I have been given a fair amount of responsibility having been sent to deal with other companies with power to purchase, alter designs and in other ways represent Atlas

Validity of the Appraisal The subsequent history of Paul Manuelli confirms, in its general outlines, the appraisal made by the counselor and is in line with the plans discussed in counseling Although the war changed the details of Paul's education, he developed in ways which had been anticipated in counseling Some of the minor, specific, ways in which his history did differ from the forecast and planning are of interest, and are taken up below

One of the counselor s misgivings, it was pointed out, concerned Paul's low spatial visualization score This had been discussed with Paul as an indication that he might do well to check his performance in drafting and related types of activities, and that trouble there might lead him to shift to a field requiring less spatial ability The subsequent history

shows no actual difficulty with spatial visualization, but the fact that his grades were not as good as the other indices would have led one to expect may be due to weakness in this special area It is perhaps significant, too, that drafting was the one subject that Paul liked least Although the stated reason was the lack of intellectual challenge, it is quite possible that the underlying reason was difficulty in transferring spatial concepts to paper There are many men as bright as Paul, with interests just as intellectual, who enjoy seeing an idea take shape on the drafting board

If it is indeed true that Paul is somewhat handicapped by low ability to visualize space relations, his future development will be worth noting, for both production and development engineering, in the mechanical field should require considerable ability of this type One might hazard the guess that Paul may well eventually "be tempted to shift to factory administration" not only because of his interest in people and planning, but also because of frustration in the more technical aspects of development work

The leadership qualities seen in Paul's high school extracurricular record and reflected in his rather high business contact interests on Strong's Blank continued to manifest themselves during his college, Navy, and business career These also suggest that, once he is firmly established as an engineer in his company, he may want to change to administrative rather than technical work On the whole, Paul's leadership record is superior to his academic record

Although it is not concerned with testing, one final point is of interest The counselor was apparently unduly pessimistic about the ability of this student to finance his way through college, pessimism which seems to have been quite unwarranted in view of the award of the four year athletic scholarship In this respect at least, the student may have shown more *savoir faire* than the counselor

Further follow-up would be highly desirable in this case, not in order to provide more guidance (Paul seems to be handling his career very well), but in order to see which predominates in the end, his technical interests and abilities or his social interests and abilities In the meantime, it may be concluded that development has been very much like that anticipated in the counseling process

Exercise 18

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what

ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

The Early Career of James G. Revere (see also p 599 and p 615)

Subsequent History The follow up of James Revere took place after a lapse of only a few months a much briefer period than in the cases of the other counselees The follow up data are therefore not very helpful for insufficient time had elapsed since counseling for the shape of events to become clear The case has been included because it illustrates, better than most recorded cases the complexity of the vocational adjustment problems presented by many men and women of about 30 years of age Briefly, Mr Revere obtained a selling job in which he thought he might use his previous training and in which he recieved training and supervision as a beginning salesman A letter from him indicated that he thought he was off to a good start in this new field

Validity of the Appraisal It is still too early at the time of writing, to judge the adequacy of the work done with this client

Exercise 19

In view of the absence of follow up material no evaluation of interpretation can be made for this case

The Early Career of Ruth Ann Desmond (see also p 601 and p 620)

Subsequent History A letter follow up for this case more than two years after counseling brought information which was as surprising as the counselor's misgivings might have led him to expect After working with the law concern as a secretary for one year Miss Desmond gave up the job and went to Pittsburgh She accepted a teaching fellowship at the University of Pittsburgh, where she taught accounting and began work toward the master's degree in business administration At the same time, she enrolled as a student of drama at the Carnegie Institute of Technology, carrying a program there which would lead to a master's degree in dramatics She completed both of these programs but insufficient time had elapsed at the time of her letter for her subsequent career to have taken shape

Validity of the Appraisal The present outcome of this case is unlike anything anticipated in the analysis of test and personal data by the

counselor or in the counseling interviews A rereading of the case (p 601 and 620) will remind the reader that the counselor had thought of statistics (machine work) and secretarial work leading to administrative responsibilities as suitable outlets for Miss Desmond, and that her employment by the law concern was appropriate The fact that she left it after about a year suggests that it was not actually so, although the reason is not clear Perhaps her poor family relations had something to do with it In the absence of detailed interview or projective test material one can only speculate

It may be more profitable to inquire whether there was something pulling her toward the field of dramatics, than to speculate as to what drove her out of secretarial work Her high literary-legal interest scores on the Strong and Kuder may have had something to do with it, moderately high artistic interest on the Strong Blank may also have played a part If her good personality inventory scores were, as suggested, compensatory in nature, a search for artistic and literary outlets for her emotions may have been another factor

Most important of all, in the writer's mind, is the uncomfortable feeling he had in working with and closing the case This feeling carried with it the conviction that there were unsolved problems in Miss Desmond's case which would keep her unsettled and on the move in search of happiness This insight, or perhaps it was only a hunch, was not the result of tests or test data, but rather of interview data and observation Whether or not it was correct can be ascertained only by her subsequent history, but that part of it which had elapsed at the time of her letter was not reassuring The carrying of two such different loads as those taken on in Pittsburgh hardly seems like a normal or well conceived plan

Exercise 20

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

The Early Career of James L Johnson (see also p 604 and p 623)

Subsequent History The hoped for opportunity to accept employment as administrative assistant to the head of a large business enterprise ma

terialized, and Mr Johnson wrote the counselor a week or so after the final interview that he had accepted the offer He was very pleased with the nature of the work, with the associates with whom it threw him, and the excellent salary which he was paid He expressed his appreciation of the counselor s services in helping him to clarify his objectives and to carry out his job hunting campaign

A follow up letter was sent to this client two and one-half years after he took the job as administrative assistant, expressing an interest in knowing how he liked the work, the nature of his subsequent experiences, and what he thought of his present situation An immediate reply was received, a brief but friendly letter in which Mr Johnson summarized the experience of the intervening two years He was still working in the same position, and had had a substantial raise at the end of his first year He felt that the prospects with his present company were excellent, and he had developed contacts through his work which might well lead to other opportunities should he be interested The work had proved to be very much to his taste detail was taken care of by a competent office force, and his own duties involved development work and contacts with a variety of executives both in the company and in other concerns There was no indication in the letter, of anxiety or difficulty in interpersonal relations such as it was thought might develop at the time of counseling While failure of such signs to appear in a letter proves little, it did seem significant that a formerly dissatisfied man wrote a letter in which satisfaction was the only manifest attitude

Validity of the Appraisal In this case, unlike Marjorie Miller's, the only follow up data are general They give blanket confirmation of the appraisal made at the time of counseling, insofar as type of activity in which success and satisfaction might be found were concerned The client's work was of a type which counselor and client had agreed should prove satisfactory, and it had proved satisfactory over a significantly long period of trial

The counselor s belief that emotional maladjustment might create difficulties for Mr Johnson did not seem to be substantiated As was pointed out, however the failure to find confirmation of this belief in a letter hardly constitutes proof The fact that the letter expresses satisfaction with the job and with its prospects may nevertheless be taken as some evidence of the fact that, as in the past, Mr Johnson was handling whatever emotional problems he might have with considerable success, winning the confidence of his employers and carrying out his work effectively

Exercise 21

Compare your interpretation of the test data and the plans which you considered suitable with the report of the subsequent history of this client In what ways were your ideas on the case borne out by experience? In what ways do you seem to have been wrong? What do you think may have been the causes of your mistakes or of the mistakes which the tests led you to make? How do discrepancies between your test interpretations and the outcomes of the case add to your understanding of the validity data reported in investigations using the tests?

Conclusions

The seven cases summarized and discussed in these chapters have served to illustrate the nature and use of data from a variety of tests, together with the need for personal data as a background against which to interpret test results They illustrate the way in which tests sometimes serve to predict with considerable accuracy the type of field in which success will be found (Stiles, Sheridan), sometimes foreshadow in a general way developments which they cannot forecast (Miller, Manuelli, Johnson), and in still other cases leave one with a baffled feeling of not having gotten to the heart of the matter (Revere, Desmond) In some cases the tests yielded important insights which could not well have been obtained by other means (Sheridan, Revere), in others they merely seemed to confirm what other data revealed (Stiles, Manuelli), and in still others they contributed to an understanding of the client but did not point the way to immediate solutions (Miller, Johnson, Desmond)

Problems of test interpretation for vocational counseling at several different levels have been illustrated Four cases were high school students one of them considering a skilled trade, three of them college majors One was a young adult with a high school education, concerned about progress in the clerical or subprofessional technical field One was a recent college graduate, dissatisfied with the occupations for which her major field had prepared her And one was a man in his mid-thirties, seeking to re-establish himself on a higher plane after the war than that at which he had worked before the war

Many more cases would be necessary in order to illustrate all the points which a user of tests should be familiar with in practice But the problems presented in this chapter, and the opportunity provided for the student to work out his own answers to them before reading what actually transpired, should provide sufficient exercise with "paper cases" It now

becomes incumbent upon the student-counselor to test some live clients, analyze the test results and relevant personal data, prepare psychometric reports in which he draws on all of the knowledge of the educational and occupational significance of tests which the study of this book and of the original studies on which it is based should have given him, and obtain the criticism of a qualified supervisor As he works with students or clients, and makes his own formal or informal follow-up studies, he will gain that richer understanding of tests of occupations, and of vocational and clinical psychology which is the earmark of the well-rounded counselor or personnel worker

APPENDIX A

STATISTICAL CONCEPTS

THIS appendix consists of two sections the first dealing with common statistical terminology, the second with the concept of prediction as applied to psychological testing in vocational guidance and selection The first section is elementary and is included only as an aid to those who may not have the background in measurement which the reading of this book and of the literature on test validities requires It does not attempt to serve as a manual of statistics or to provide sets of statistical tables Those are available in Garrett (282) and Walker (906) It should be skipped by all who are familiar with the terminology and concepts of statistics The second section is of more general interest and contains material important to many readers who have some knowledge of statistics It also emphasizes logic rather than formulae or tables (see p 654)

STATISTICS QUANTIFIED REASONING

Those who are not used to working with numbers or who have had inadequate instruction in mathematics in high school, often approach statistics and even reports which include statistically presented data with their minds made up that they cannot understand them Such readers should bear two things in mind First as shown in Chapter 6 the relationship between verbal ability and achievement in mathematics at the senior high school and college levels is as close as that between quantitative aptitude and mathematics therefore an intelligent person can learn what he needs to in the way of high school or elementary college mathematics Second statistics is nothing more than logic expressed in numerical form therefore any reader who can engage in logical reasoning can master elementary statistics and those who enjoy logic should enjoy statistics

It is not the purpose of this section to convey any knowledge of statistical formulae and computation That is not necessary for the reading or understanding of this book But as an understanding of the *concepts* of statistics is necessary both for the reading of books such as this and for the interpretation of tests, the following paragraphs attempt to explain briefly the meaning of the commonly used statistical terms

Central Tendency

Tests are measuring instruments Measurement generally involves the comparison of one entity with some other entity This may be the weight of a person and of some pieces of iron, the length of a table and of a standard-sized object called an inch (in French an inch is a "pouce, or thumb), or the number of words understood by a first grader and the number understood by other first graders In psychological work the comparison of a person with something usually requires that he be compared with other persons Although one can simply count the number of words understood by a first grader, that number has little meaning unless one also knows how many are understood by the average first grader

In order to make such comparisons, the typical achievements, aptitudes, interests, or personality traits of various kinds of groups must be expressed in summary form Numerical summarizations of group characteristics are called measures of central tendency Measures of central tendency are averages Averages can be expressed in several ways, as medians, means or modes

Median The median (Md) is the middle person or middle score in a distribution of persons or scores In a group of five boys standing in the order of their height the third boy from either end is the median boy This method of expressing averages is commonly used when the number of cases is not large, when a few extreme cases might distort other indices of central tendency and when a quick estimate is desired

Mean The mean (M) is what is most often meant when, in everyday language, we talk about averages It is computed by adding up the ages, heights, or I Q s of all persons in the group and dividing the sum by the total number of cases The mean is the most widely used statistical measure of central tendency because it is part of a system which lends itself to many types of manipulation When the number of cases is small, however it can be seriously distorted by a few extreme cases

Mode The mode is the height age, or test score which is most common in the group In a perfectly normal distribution it is identical with the median and mean but in skewed distributions it is not the same It is ascertained by inspection, to see where most cases fall A distribution may be bimodal, or have two modes, in very special cases but it would still have only one mean and only one median The mode is rarely used however, because it does not lend itself to as wide a variety of applications as the mean or even the median

Dispersion

In order to describe the status of a group one must know not only what is typical, but also the extent to which the group varies from its own norm One company s salesmen may be a very homogeneous group in so far as intelligence test scores are concerned, most of them having I Q s very close to the mean of

110, in another company the mean may be almost identical, let us say 111, but the variability greater, some making much lower scores and others much higher Despite comparable means, the two groups are clearly different in intelligence Measures of dispersion are used to describe the extent to which the cases cluster around the average or scatter away from it These include the range, interquartile range, and standard deviation, plus some others which are less widely used

Range The range is the simplest and crudest measure of dispersion It is the difference between the lowest and the highest cases Thus the ranges of I Q s in the groups of salesmen just mentioned may be from 105 to 116 in the first group, and from 95 to 121 in the second It is not a good measure of variability, because a few extreme cases may give the appearance of considerable dispersion when most cases actually cluster close to the median

Interquartile Range The interquartile range includes the middle fifty percent of the group By telling how far out from the average this half of the group spreads, it gives a reasonably good idea of how representative the average is of the group as a whole The semi interquartile range, the distance including 25 percent of the cases on one side of the median, is also sometimes used These measures are used with the median, and are part of the *percentile system* Another way of describing the interquartile range is to say that it extends from the 25th percentile to the 75th It is computed by finding the test score which is made by the person who is one fourth of the way up from the bottom of the distribution of scores, and that which is made by the person who is three fourths of the way up These two scores or points on the distribution are called the first and third quartiles (Q) The median is of course the second quartile, and the high end of the range is the fourth The term interquartile range is not often used, but Q_1 and Q_3, the first and third quartiles are generally used with the median in order to describe the variability of the group

Standard Deviation The standard deviation (sigma or σ) is the measure of dispersion commonly used to describe the variability of groups for which the means have been ascertained It is virtually an average of the distances of all the scores in a distribution from their own average score Means and standard deviations are part of the *moment system*, just as medians and quartiles are part of the percentile system The standard deviation is, with the mean, the more commonly used measure because it lends itself most readily to use in other formulas The distance between one sigma either side of the mean of a normal distribution includes, not 50 percent of the cases as in the interquartile range, but the middle 68 percent This number may seem awkward but there is no special virtue in 50 percent, and the standard deviation actually gives a somewhat truer picture of the scattering of cases or scores around the mean

One fundamental difference between the percentile and moment systems should be kept in mind percentile, quartile and other such scores are based

on the *number of cases,* but standard deviations or sigmas are based on *distances from the mean.* The latter are therefore more truly measuring sticks in the usual sense of the term and give a better idea of dispersion than do the former.

Methods of Expressing Scores

Test scores are commonly expressed as raw, percentile, or standard scores. The raw score is simply the number of problems correctly solved, the number of words known, or some other index of work done. It therefore needs no special defining. But raw scores are not meaningful until they are converted into some other type of score which shows the examinee's standing with regard to a group of persons in the same occupation, school grade, or problem group. Percentile scores and standard scores have their own special virtues and defects, which must be understood to be wisely used.

Percentile Scores or Ranks. As was indicated above, percentile ranks are based on the frequency with which cases fall at given points on the scale. They have the advantage of being based on a concept which is familiar to educators and to people in general, who readily grasp the significance of the statement that a college senior has more mechanical comprehension than 80 out of 100 applicants for engineering positions or more ability to perceive differences in pairs of numbers than only 10 out of 100 clerical workers. This is what a percentile score tells one when cases are counted from the bottom upward (the usual method). But just because it is based on counting cases this system has the defect of making differences near the median seem greater than they are, and of disguising differences at the extremes. To illustrate this latter defect, it is necessary only to point out that two people, one with an I. Q. of 190 and the other with an I. Q. of 180, are both at the 99th percentile. Both are more intelligent than 99 out of 100 persons, but the difference in their mental ability is very great. If one used only the percentile score, as is done with most aptitude tests, the difference would be hidden.

Standard Scores. Standard scores, being based on distances from the mean, provide sensitive indices of abilities and traits. Most systems arbitrarily assign a standard score of 50 to the mean raw score, and make 10 standard score points the equivalent of one standard deviation in raw scores. Thus if the mean raw score in a normal distribution is 124, the mean standard score is arbitrarily called 50. If the standard deviation (the distance either side of the mean which includes 68 percent of the cases) of these raw scores is 40, then 124 (the mean raw score) plus 40 (the standard deviation in raw score points) equals 164, which is a standard score of 60. If two sigmas in raw score points were added to the mean raw score, one would have 124 plus 80, or 204; this would equal a standard score of 70. The mean raw score minus one sigma (124–40) equals 84, which is a standard score of 40. Minus two sigma would be a raw score of 44, or a standard score of 30. Most actual standard scores are between 30 and 70.

Some standard score systems differ slightly in method of expressing scores but not in logic T scores are sometimes the same as those described here, but sometimes consist of 100 units of 1 sigma each ranging from −50 to 50 Sigma scores are the same as standard scores except that they use only one digit to the left of the decimal point, and therefore one or two to the right (mean equals 5 0) It is possible also to designate the mean as zero, and to use true sigma scores, showing scores as positive or negative (one sigma above the mean would be 1 0, one below −1 0) Army standard scores, used in the Army General Classification Test, have a mean of 100 and a standard deviation of 20, and range from 40 to 160, three sigma below and above the mean

The Significance of a Difference

When the test scores of two different groups are being compared, it is important not only to have measures of central tendency and dispersion, but also of the significance of whatever differences are found between these measures One must ask not only whether the mean score of one group is higher than that of the other, but also whether it is sufficiently higher for one to have some confidence that future samples of the same populations will differ from each other in the same way In asking these questions one passes from descriptive statistics to the statistics of inference Instead of describing the status of a group, one generalizes from a known group to other similar but unobserved groups An objective answer to this question is provided by measures of the significance of a difference, the most common of which is the critical ratio (the t test)

The Critical Ratio The most common method of determining the likelihood that the difference between two groups is reliable is to divide the difference between the mean scores of the two groups by the standard error of the difference between the means of the two groups The resulting statistic is called the critical ratio (C R) This can in turn be converted into an expression of probability (p), that is a statement concerning the number of times in 100 that the obtained difference might be found strictly as a result of chance This procedure is also known as the t-test When the number of cases exceeds 30, a critical ratio of 2 00 means that there are 5 chances in 100 that a difference such as that obtained between the two groups would be found in a situation in which there were really no differences A critical ratio of 2 50 shows that there is only one chance in 100 that the difference was due to chance factors A critical ratio of 3 00 means virtual certainty that the observed difference is a real difference, for chance could produce a difference of that order only three times in 1000

Decisions as to what actually constitutes a significant difference vary partly with the conservatism of the judge and, more legitimately, with the nature of the decisions to be made A man out for a stroll might well avoid a bridge if there were 5 chances in 100 that it would collapse under his weight for he could just as well walk elsewhere with less risk But he might gladly cross it with only a

50–50 chance of its supporting his weight if the safety of his small child depended upon his doing it

Relationship

In order to understand what a test measures, one must know what scores on that test are related to, and the degree of that relationship If a test is supposed to measure understanding of mechanical principles and processes, there should be available evidence that it is related to success in mechanical activities The common measures of relationship are coefficients of correlation There are a variety of these, using some form of the letter r as symbol product moment or zero order, rank order, biserial, tetrachoric partial and multiple correlation coefficients There are other measures of relationship such as the coefficient of contingency (C) and the correlation ratio (eta) but as these are not commonly used in analyzing test data they are not discussed here

Product Moment Correlation When the results of two measures are expressed in terms of scales they can be related to each other by the product moment or Pearson correlation coefficient (r) This is the common method of determining such things as the extent to which intelligence and school marks vary with each other and the degree of association between sales interest and success in selling life insurance These correlations can be plotted graphically, as in Figure 26

The scale on the left hand side of the graph shows intelligence test scores (I Q s) the higher the score on the scale, the more intelligent the individual The scale on the base line shows occupational level the further to the right one goes, the higher the person s standing on the occupational ladder An individual who makes an intelligence quotient of 110 and an occupational level score of 75 is represented by a stroke in the cell or box located at the intersection of the perpendiculars (broken lines) leading from the points corresponding to his scores on the side and base lines Inspection of these strokes shows that, the higher the location of an individual on the intelligence scale (left hand side), the higher (more to the right) his status is likely to be on the occupational scale (base line) A correlation coefficient is nothing more than a quantitative and therefore objective method of expressing the extent to which these strokes fall in or deviate from a straight line drawn from the lower left hand corner to the upper right hand corner of this graph If all of the strokes fell on a straight line, one could tell exactly what the occupational level of an individual is from a knowledge of his intelligence test score The coefficient of correlation would then be 1 00 showing a perfect relationship If the strokes scatter around this line the relationship is positive but less than perfect, as shown by the extent of scattering away from the line and by correlation coefficients ranging from 20 or 30 to 90 or something less than 1 00 Sometimes the relationship is negative then the diagonal in Figure 26 runs from top left to bottom right and coefficients of — 20 or — 30 to something less than —1 00 would be preceded by minus signs

to show that the higher a person's standing on one measure, the lower he is likely to be on the other This might be the case, for example, with intelligence and number of bookkeeping errors

Correlations of 0 00 to 20 generally indicate a lack of relationship between two measures, whatever the sign But how large a correlation must be to be

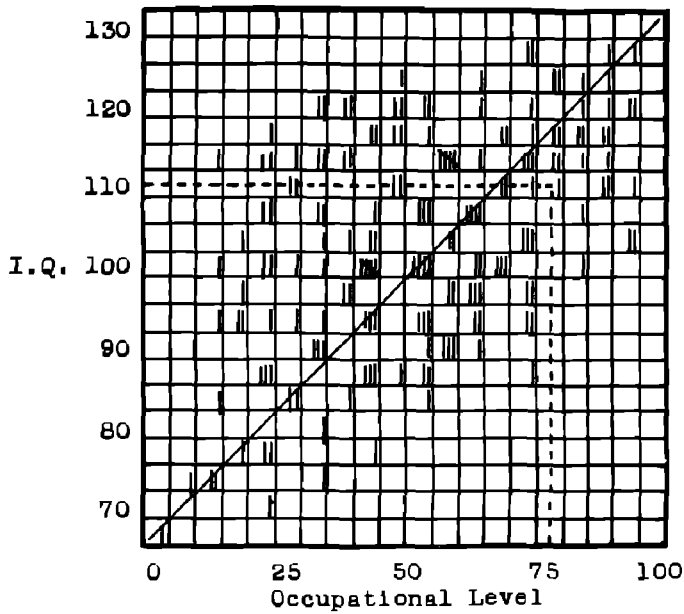


FIGURE 26

SCATTERDIAGRAM

Relationship Between Two Traits

Each vertical line in the scatterdiagram represents *one case* Thus the single stroke at the junction of the two broken lines represents one person with an I Q of 110 and an occupational level score of 75 The closer the strokes (cases) to the diagonal line (lower left to upper right) the higher the correlation between the two traits the more they scatter away from it, the lower the correlation

significant depends upon the number of cases involved and upon the reliability of the measures For this reason the *probable error* of a correlation coefficient is often appended by means of plus and minus signs (e g , 24 ± 07) If the probable error is as much as one fourth the size of the correlation the obtained relationship may be due to chance factors As the probable errors of correlation coefficients tend to be as high as 05 or 08, the correlation generally has to be above 20 or 30 to be statistically significant Even then it may not be practically

significant for, as will be seen later, a correlation of .30 improves the efficiency of a prediction by only 5 percent above chance Some investigators (the statistical purists) prefer to report the significance of a correlation in terms of the probability of its occurrence by chance in such cases it may be stated that, for a relationship in a group of such a size to be significant at the one percent level of confidence it would have to be .35 (or some other figure) If the obtained correlation is lower, it cannot then be said to be statistically significant or high enough to be genuine The *level of confidence* statement is one of probability the 5 percent level of confidence means, for example that such a relationship would occur by chance only 5 times in 100 As in the case of the significance of a difference the user of the test must then decide whether the degree of confidence which he can have in the obtained relationship is great enough to serve as a basis for making the kind of decision which is being considered *This depends also upon the degree of confidence which he can have in alternative bases, which is too often not ascertained* Assuming large enough numbers and low enough probable errors correlation coefficients are generally defined in the following terms

.80 and up		very high correlation
.50 to	.80	substantial correlation
.30 to	.50	some correlation
.20 to	.30	slight correlation
.00 to	.20	practically no correlation

A word of caution should be inserted here concerning the meaning of the terms relationship and correlation There is a very common and natural tendency to translate these into cause and effect But relationship means, statistically as well as in everyday language, that things tend to be associated, to go together Cousins are related but they are not one the cause of the other It is true that common sense tells us that intelligence causes good marks in school, and that the two do not merely happen to go together or even to be the joint effects of a common cause But that is what common sense tells one not correlation statistics All the correlation coefficient shows is that students who are intelligent (or who get good marks) tend also to get good marks (or to be intelligent)

Rank-Order Correlation This is another method of computing relationship, simpler than the product moment method and superior when only a few cases are involved It requires only that they be ranked in order of standing on the tests or other measures Logically the concept is the same as for the more accurate method just discussed In a perfect rank-order correlation (rho), for example the student or employee who stood first on one measure would also stand first on the other and so on down the line if the relationship were negative, the highest person on one would be the lowest on the other, the second highest on one second lowest on the other, etc

Biserial Correlation Sometimes one of the variables being analyzed cannot be expressed in terms of scores or numbers on a scale Thus the criterion of success may be ability to learn to fly an airplane, keeping vs losing a job, or some other indication of ability or status which has only two categories In that case the biserial coefficient of correlation (rbis) is used It is interpreted in the same manner as Pearson and rank order correlations

Tetrachoric Correlation This type (rtet) is used when both of the measures are dichotomous It is therefore rarely encountered in studies of tests

Partial Correlation It is sometimes necessary to ascertain the relationship between two measures when the influence of a third variable is held constant or ruled out For example, if it is desired to find out the relationship between perceptual speed and success in machine bookkeeping and if intelligence affects both scores on clerical perception tests and success in machine bookkeeping the correlation between clerical perception scores and bookkeeping success will seem unduly high The common third factor will make it so Partial correlation ($r_{12.3}$) techniques makes it possible to hold constant or eliminate the influence of other measured factors, such as intelligence in this example by statistical methods The interpretation of partial correlation coefficients is similar to that of Pearson r s

Multiple Correlation When more than one test is used, as in selection batteries it is necessary to ascertain the relationship between scores on the combination of tests and the criterion of success This is done by multiple correlation (R), in which product moment correlations for each pair of variables are first computed separately and then combined The interpretation of R is similar to that of r

Reliability One fundamental question which needs to be answered for every test has to do with the extent to which it agrees with itself If it gives the same results both times when used twice with the same person or if a score based on one half of the test agrees with a score computed from the other half it can be used with confidence that it is measuring something and measuring it consistently This is known as reliability and is expressed as a correlation coefficient If on the other hand, the test does not agree with itself when repeated (retest reliability) or divided into halves (split half or odd even reliability), one cannot even be sure that the test is measuring something Of course some variation in scores is permissible because of chance factors practice effects etc, but the reliability of a test for use with individuals should be 85 or above It should be noted that the fact that a test is reliable (measures something consistently) does not prove that it is good for anything This latter question is one, not of reliability, but of validity

Validity The second major application of correlation statistics to testing is in the determination of the extent to which a test measures that which it purports to measure or, in fact, anything else that the test user thinks it might be

desirable to measure A test devised to measure a type of auditory acuity deemed to be important in music was found also to be important in submarine detection work (see p 321), a test designed to measure aptitude for mechanical activities was useful in selecting typists (page 278), and both of these were valid also for the purposes for which they were designed There are, on the other hand, many published tests with names which imply that they are valid for some special purpose but with little in the way of evidence to prove the implication

Evidence of the validity of a test is most often presented in the form of a correlation coefficient which indicates the degree of relationship between scores on the test and some external criterion such as grades, ratings of supervisors, earnings output, or job satisfaction When the use of common correlation techniques is not possible (e g when the criterion is not scaled but consists only of distinct categories or of two classes such as successes and failures), other less refined measures of relationship are used Some of these should actually not be called measures, as they do not indicate the *degree* of relationship but only the fact that a relationship exists They are more appropriately called tests of the *significance* of a relationship The chi square test is one of these their interpretation need not be discussed here, as they are always finally expressed in terms of the probability that a comparable relationship might be found on a chance basis

The phrases *internal* and *external* validity are frequently encountered in the literature, and the concepts are met even more often in test manuals A test author's logic in selecting the content of his test has been clear cut and stringent each item in the test has a high correlation with the total score (it has internal consistency), and the test is reliable these and other such facts are commonly cited by authors of new tests as evidence of the validity of their instruments Such evidence is called internal evidence of validity, as it is based entirely on analysis of the content of the test without reference to objective external criteria Reliability indicates objectively that the test agrees with itself but does not tell what it is that it measures Internal consistency is another aspect of the same thing And a critic's evaluation of the appropriateness of the content of the test involves only subjective reference to external evidence, which is the same as that used by the test author in devising the items author and critic could therefore be making the same judgmental errors Therefore test manuals which cite only internal evidence of validity say, in effect, *"Caveat emptor"* The warning should always be made explicit, and publication of such a test should carry with it responsibility for taking the next steps

External evidence of validity is, then the only type which really provides an adequate basis for judging a test, and tests lacking it are suitable only for experimental use Types of criteria against which a test can be validated are discussed in Chapter 9 of this book It is brought out there that finding appropriate and usable criteria is by no means simple As they consist of such things as

ratings, output, and other such variables they usually lend themselves to the use of correlation techniques

The minimum acceptable for a psychological test has generally been set at .45 This figure is selected because a prediction on the basis of a test or battery of tests with this degree of relationship to the criterion would be 11 percent better than a prediction based upon chance to put it another way, predictions based on such data would be correct in about 55 out of 100 cases wrong in about 45 out of 100 (that the figure .45 occurs twice in this context is not an indication that correlations can be treated as percentages but is due to other factors which need not be gone into here) The setting of a minimum acceptable validity coefficient, whether of .45 or some other figure, has had the unfortunate effect of making many people conclude that a test with less validity for a given purpose is therefore of no value This involves a logical fallacy which should be cleared up

It is true, as the data imply, that a relationship expressed by a validity coefficient of less than .45 is so slight as to be of little practical value *by itself* The fallacy is the assumption that it is used by itself In practice test data are subjectively combined with other data in estimating probabilities, whether in counseling or in selection These other data may consist of evidence of financial backing which will make possible an educational or a business venture judgments of motivation and drive amount and type of education received, etc Each of these also generally has relatively little relationship with success but the counselor or personnel manager trusts that by depending on a combination of such considerations he will make better judgments than would otherwise be made

What psychology and statistics do is change trust to probability and convert judgments to measures A comprehensive test battery is a series of measures of educational background socio economic level intelligence and whatever else is related to success in the occupation in question Each of them is known to be related to an appropriate criterion of success as shown by a validity coefficient They are combined, not by the judgment of an individual but by a regression equation which gives to each variable the weight which experience has proved it should have

Experience with batteries of well constructed and varied tests has shown that measures with validity coefficients as low as .20 may be useful (provided the correlation is statistically significant) It is true that, if such a test were used alone, the predictions would be right only 51 times out of 100 But if this test measures some trait or aptitude which is unrelated to other factors measured by a battery of tests, it will add appreciably to the validity of the battery An illustration of this fact is discussed in connection with the development of a custom built personality inventory for pilots (see pp 528 ff) In this investigation a test with a validity coefficient of .20 and a low correlation with the battery raised the validity

of the battery from .66 to .70. This improvement in the correlation between tests and criterion would have resulted in the selection procedure being right 65 percent instead of 63 percent of the time. The gain was relatively slight, but was made at a cost of only 20 minutes of testing and less than one minute of scoring, at a stage when finding any tests which improved the battery at all was extremely difficult.

This brings out a final point concerning validity coefficients: they are not likely appreciably to exceed .70, because of the unreliability of criteria. The logic of this should be clear if it is remembered that when two supervisors rate the same employee their ratings do not agree perfectly, or that when two teachers grade essay examinations the grades they give are by no means identical. If the two sets of ratings are thought of as two measurements of the same thing (which is what they are intended to be) then it is clear that the coefficient of correlation which expresses the relationship between the two sets of ratings is a reliability coefficient. The ratings of two supervisors infrequently reach an intercorrelation of .70, falling short of the desired reliability of .85 or better. When the criterion agrees so poorly with itself, one cannot expect even a test with a reliability of .95 to correlate very highly with the relatively unreliable criterion.

In summary then, tests with validity coefficients of as little as .20 may be useful; the combined predictors (some of which may not be tests) should have a validity of .45 or better to be appreciably better than chance, and no combination of tests is likely to yield a validity coefficient much above .70.

Prediction and Probability

The use of the term "prediction" in the literature of vocational psychology has been widespread. Thorndike's early study (1928) was entitled *The Prediction of Vocational Success*, and a much more recent book sponsored by the Social Science Research Council and edited by Horst (1941) is entitled *The Prediction of Personal Adjustment*. The articles in professional journals which use the term are legion. The result is an impression that prediction is one of the major functions of applied psychology.

Some Misgivings About Prediction

But the term prediction needs to be defined, and the type of prediction undertaken by vocational psychologists and counselors needs to be made clear. Kitson (1921) has forcefully expressed the misgivings of many psychologists concerning the use of this term. "Once we recognize the influence of any or all of these [personal and situational] factors on the vocational success of an individual, we must acknowledge how futile and presumptuous it is to administer a few tests to an individual and, from his scores, to attempt to foretell his eventual success or failure. . . . Optimistic psychologists sometimes declare that we shall be able to predict vocational success 'when vocational tests are more highly developed.'

On this point, William James made a pertinent observation sixty years ago 'It is safe to say that individual histories and biographies will never be written in advance no matter how evolved psychology may become

Allport has voiced similar misgivings (12) 'The fact that 72% of the men having the same antecedent record as John will make good is merely an actuarial statement It tells us nothing about John If we knew John sufficiently well, we might say not that he had a 72% chance of making good, but that he, as an individual, was almost certain to succeed or else to fail

Multiplicity of Factors

Underlying the misgivings of writers such as Kitson and Allport is the recognition of the fact that a person's actions are determined by a great variety of forces some of them residing within the individual some of them essentially part of the environment Horst (383 13–18) has discussed these in some detail

Personal factors may be either congenital or environmental in origin As has been well brought out by a number of investigations (e g , 568) both constitution and environment play some part and it is difficult to untangle their relative importance This fact is of significance to users of psychological tests in guidance and selection, because it makes their task that much more complex the modifiability of a trait or aptitude makes it necessary not only to know the chances of a person with a given amount of it succeeding in a given activity, but also the chances of his having intervening experiences which modify the degree to which he possesses the trait (not to mention the probability that having experience in the activity itself will modify it)

Situational factors are often mensurable, but in many cases cannot be assessed The latter are often referred to as chance factors or luck Among the situational factors affecting success which can be measured and controlled are differences in the purchasing power of sales territories affecting the production of salesmen, differences in the aspiration levels of cultural groups affecting the output of factory workers and the possession of private pilot's licenses by members of the family of would be aviation cadets, affecting their motivation to fly and their orientation to flying Some typical unpredictable situational factors which affect success in vocational endeavors are illness in the family, which makes the individual geographically immobile and drains energy which might otherwise go into his work, atmospheric conditions which make bombing difficult for bombardiers in that locality, the colleagues with whom the individual must work, such as a *dishonest partner or a selfish collaborator and the outbreak of war,* which handicaps persons in some occupations and materially aids those in others

It has been pointed out by Horst (383 55) that one of the chief reasons why many prediction procedures have not attained a higher level of accuracy has been *their failure to take into account contingency factors Contingency factors* are those personal and situational factors which affect performance but for which the

probability of subsequent presence or absence is not known at the time of prediction Thus there is no way of knowing what the health of unborn children will be, with its possible effects on the occupational mobility of the father, nor is there any way of knowing when he is a sophomore in college the particular type of sales job and territory a potential salesman will get It is the failure of most psychologists who write about 'prediction' and publish studies of the predictive value of tests to take such factors into account that has led Allport, James, and Kitson to criticize the use of the term and to despair of actuarial predictions in applied psychology

Taking Contingency Factors Into Account

But others have not been so pessimistic, as is brought out by the mere fact of the publication of the Social Science Research Council's monograph (383) Horst lists three methods of dealing with contingency or "chance" factors which have been proved promising

1 *Adjust the criterion score in terms of the contingency* Thus of two salesmen with equal sales volume but in territories, one of which has a high level of purchasing power and the other a low level the salesman in the latter is to be considered more successful and his criterion score (sales volume) is corrected by a statistically derived weight This shows that he has really been more successful than his mate in selling an equal amount under more difficult conditions

2 *Treat the contingency factor as one of the predictive elements* In the case of the potential salesmen still in college this method would not work, for there is no way of knowing what type of territory he will be given Although suggested by Horst this technique is actually not applicable to true contingency factors, for if by definition the probability of their presence cannot be known at the time of prediction that item cannot be scored in making the prediction Horst's example (383 56) is drawn from the prediction of academic success, he states that weights may be assigned for given amounts of time spent in outside work But this involves knowing how much time is spent in outside work by a given individual The so called contingency factor then becomes a known variable and a predictive factor The procedure is comparable to scoring a would be life insurance salesman's biographical data blank according to the known relationship of age marital status and amount of insurance carried Prediction studies in which such ascertainable factors are not included among the predictors are legitimately to be criticized, if prediction is to be attempted all potentially revelant and mensurable factors should be included among the predictors

3 *Predict the contingency factor* If the college student who is considering a career in life insurance is to be tested and an estimate of his probable success in that field is to be attempted it is difficult to know how to weigh such items as marital status and amount of insurance carried At his present age he has not had the opportunity to marry or to carry insurance which he will have had after he

has been out of college several years But the probability of his getting married and carrying insurance can be ascertained Horst suggests that the prediction formula might include factors related to subsequent marriage and that in predicting the contingency (marriage) the prediction of the activity (selling) can be improved

The application of these methods adds considerably to the complexity of the prediction procedure In the case of military pilots in World War II for example, it meant that the prediction of success of fighter pilots in combat had to be broken down into success as a fighter pilot with equipment superior to that of the enemy in a theater with considerable air opposition as a fighter pilot with equipment superior to that of the enemy in a theater with little air opposition as a fighter pilot with equipment inferior to that of the enemy in a theater with considerable air-opposition and other such categories It meant that the test battery had to include not only aptitude tests of the usual types but also biographical data blanks covering such factors as marital status and age (younger men are more likely to succeed than older but married men are better risks than single) previous flying experience (those who flew voluntarily as civilians are most interested) having a pilot relative (having a flier in the family seemed to mean having flying in the blood) and urban vs rural origin (city boys are less likely to succeed in flying training than those who are more used to outdoor life)

Even when the criterion becomes as refined as possible and even when the list of predictors is made as inclusive as job analysis man analysis and ingenuity of test construction permit there are still many factors which are not covered These are the true contingency factors those which are most truly matters of chance such as the honesty of a partner the outbreak of war and epidemics

Probability

In view of the fact that no prediction of human behavior vocational or otherwise can take into account all relevant factors it seems wise to use the term prediction cautiously and with a full awareness of its definition As used by statisticians the term predict is more or less synonymous with to estimate , as in the prediction or estimation of weight from height or of a son s intelligence from his father s Knowing one it is possible to make a best estimate of the other which while often not accurate is much better than a pure guess There are times when one or more such correlates are known when others are not known and when decisions or judgments need to be made The best estimate is then helpful

But a best estimate is merely a statement of probability It says in effect there are 7 chances in 10 that this man is not heavy enough to move this load intelligent enough to succeed in a highly selective college or aggressive enough to make good as a house to house salesman whichever the case may be It should be noted that these statements are not *predictions* they are statements of the

probability of one specific type of behavior in one specific type of situation It is not success in factory work, in college, or in sales work which is predicted Rather, it is the probability of a person being heavy, bright, or aggressive enough to perform a specified task which is estimated The form in which the estimate is expressed makes it clear that other factors which may affect success are not

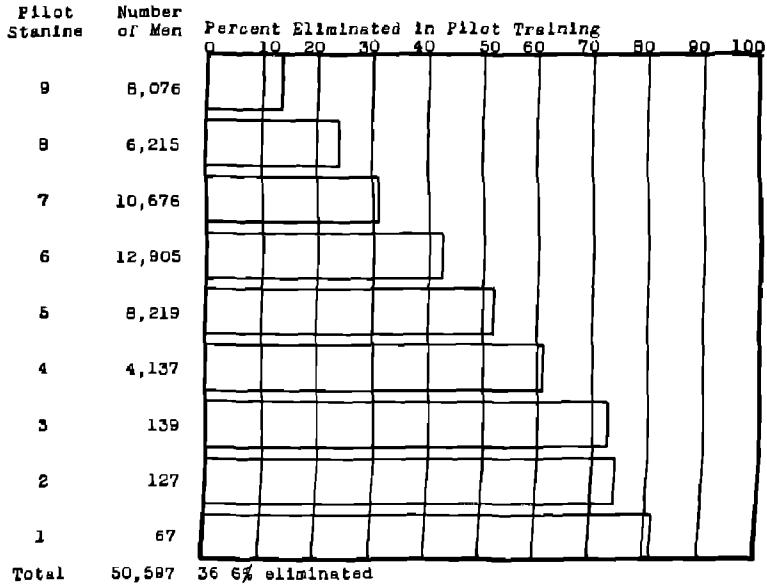


FIGURE 27

DATA FOR ESTIMATING CHANCES OF SUCCESS IN PILOT TRAINING

Actuarial data for this group of more than 50 000 aviation cadets showed that cadets with stanines of 9 have 87 chances in 100 of completing pilot training (primary through advanced) whereas those with stanines of 5 have slightly less than a 50–50 chance of completing training and only about 19 in 100 of those with stanines of 1 succeed After DuBois (214 145)

taken into account The estimate of ability to move the load, obtain passing grades in college or make sales can be made still better by taking other things into account, this may be done as Horst suggested, by including them among the predictors or in the criterion, or by subjective modification of the probability estimate For example one might take into account the previous physical activities of the laborer and the type of equipment used to move the load the educational achievement of the college student's mother and his own expressed

attitudes toward college, and the financial need and past social achievements of the salesman

Estimating the Probability of Success in Flying Perhaps the meaning of estimates of probability in vocational guidance and selection can best be made clear by means of a specific example 50 000 odd cadet pilots processed by the AAF aviation psychology program Figure 27 shows the number of men in each stanine group, and the percentage of failures in each group The length of each bar is proportionate to the percentage of men eliminated from flying training Approximately 81 percent of the men who were sent to flying training despite stanines of 1 (a practice discontinued early in the war) failed to make good More than 70 percent of those with stanines of 2 and 3 also failed But only about 13 percent of those who made stanines of 9 and 24 percent of those with stanines of 8, failed to make good As shown also on page 23 for all stages of training this battery of tests obviously had considerable value in differentiating the men who were likely to succeed from those who were likely to fail in flying training The relationship in Figure 27 is expressed by a biserial correlation coefficient of 38 between pilot stanine and success in pilot training which is raised to 49 when corrected for restriction of range (it was 63 for an unselected experimental group of over 1000 cadets [214 191])

It is pertinent to point out that the battery of tests used included tests such as those which constitute intelligence tests (arithmetic reasoning and reading comprehension) tests of spatial visualization, general information, mechanical comprehension mathematical achievement coordination, finger dexterity, perceptual speed and reaction time etc It also included a biographical data blank covering family background education occupational experience hobbies urban-rural experience, etc Although it did not include such measures of personality as the Rorschach interview impressions and the like such indices had been tried and were found to have no validity for predicting success in flying It was therefore about as comprehensive a battery as could be devised Such contingency factors as were not provided for were probably not such as could be taken care of without an undue additional expenditure of time and money With this in mind let us consider the estimates of probability the predictions," which may be made on the basis of this combination of predictors

Approximately 81 percent of the 67 pilot cadets who made stanines of 1, but were sent to training nevertheless failed to complete pilot training having been eliminated for flying deficiency or fear or at their own request The odds may therefore be said to be four to one against a person with a stanine of 1 succeeding in flying school This is a statement of probability which is rather impressive and when other candidates are available is certainly evidence in favor of not selecting those with such scores But suppose that one is concerned, not with the selection of large numbers of men from a much larger pool but rather with the evaluation of the chances that a particular individual John Smith who made

a stanine of one, will make good as an Air Force flier The odds are still four to one against him but there is conversely one chance in five that he will succeed These are not hopeless odds and Smith will certainly argue if given an opportunity, that he is the one poor risk in five who will succeed! The personnel worker, psychologist, or counselor cannot deny his contention All he can do is point out that each of the other poor risks feels the same way (they did when the writer interviewed large numbers of them early in World War II) and that experience shows that approximately four fifths of them fail nonetheless He must recognize that the *prediction* is for a *group* four fifths of it will fail For any given member of the group all one has is a *probability statement* the odds are four to one against him Only experience can show whether John Smith would be one of the 81 failures or one of the 19 successes in every 100 men like him

The same can be said of the high stanine men Of those who make stanines of 9 (8 076 men in this group of 50 597) only about 13 in each 100 fail in flying training The odds are therefore overwhelmingly in favor of the cadet who makes a score of 9 they are about 7 to 1 But 13 in every 100 such cadets did fail, and Cadet Jack Doe who made a score of 9 has no way of knowing whether he is one of the 87 or one of the 13 Neither has the personnel officer, the counselor, or the psychologist who helped develop the tests

The examples just cited are the most clear cut possible for they are selected from the extremes of the distribution Consider the men who made average scores Cadet Jim Dale for example, with a stanine of 5 In this sample of 50,000 odd who went to pilot training after taking the cadet tests there were some 8 000 who made stanines of 5 About 48 percent of these failed in flying school The odds are therefore about 50–50 that Dale will succeed, but there is no way of knowing whether he will be one of the 52 in 100 who pass, or one of the 48 who fail He may consider it worth the risk and so may others if there is a manpower shortage But when other goals are equally attractive the candidate may well prefer greater odds in his favor, and when more promising candidates are available personnel may legitimately reject him in favor of others In neither case is there any prediction that Dale will either succeed or fail there is only a statement of probability

Confusion in Counseling Unfortunately, probability statements are viewed by many persons as predictions The result is that having heard a great deal about the predictive value of tests used in selecting groups of men and women for military or industrial assignments many people come to vocational counselors and psychologists to be tested in order to find out what they are "best fitted for" The success of vocational appraisal procedures for prediction in one sense (the tendency of groups to succeed or fail) has created impossible expectations of these same procedures when used for another purpose (the appraisal of individuals) The result is a feeling of disappointment on the part of those seeking

guidance and one of frustration at not being able to work effectively, in ways appropriate to their tools, on the part of counselors The situation would be improved if the general public, and some psychologists whose absorption in test construction has caused them to lose sight of the context in which they work, would cease to think in terms of predicting the success or failure of individuals, and come to think in terms of probabilities, some of the contingent factors of which will remain unknown even after the most thorough of testing and interviewing procedures

The Accuracy of Estimates

One final point needs to be made concerning the estimation of a person's standing on one scale, let us say production records from his standing on another scale, such as a battery of tests The imperfect relationship between test battery and criterion means that instead of yielding a *point* on the criterion scale the correlation coefficient yields a *zone of approximation* Stated in nonstatistical and concrete terms when we estimate the amount of insurance which an applicant for a job is insurance salesmen will probably sell from the scores which he makes on selection tests, the result must be expressed, not as $100,000 00, but as "$100 000 00 plus or minus $90 000 00 or as "from $70 000 00 to $130,000 00 The higher the validity coefficient, the narrower the zone of approximation that is the closer the estimate is to a specific figure Conversely, the lower the validity coefficient the wider the zone of approximation or the greater the range of sales which may be made by the potential salesman When the correlation between tests and sales is zero the zone of approximation ranges from nothing to infinity what the salesman will sell is a matter of guesswork

Table 35, commonly called a prediction table makes it possible to ascertain the most probable criterion score and the zone of approximation for any given test score, given a validity coefficient test scores which can be expressed in terms of standard scores or percentiles and criterion scores which can be expressed in the same way Test scores are normally available in this form and many criterion data can be converted into these terms For example if the criterion is dollar value of insurance sold per annum, this figure can be ascertained for each salesman, and the salesman's standing can be compared to that of other salesmen of the same product for the same company, and converted into a standard score or percentile by the usual methods

To use this table, one enters it with the score on the known measure (test) by means of the score scale at the top Following the appropriate column down, one stops at the row opposite the r corresponding to the actual correlation between predictor and criterion The figure where column and row meet is the most probable criterion score The column headed h (standard error of estimate) indicates the amount to be added to and subtracted (after multiplication by 10 to match the standard scores) from the criterion score to give the zone of ap

TABLE 35

ESTIMATION TABLE

To estimate a person's most probable standing on a criterion score (expressed as a standard score or percentile) from his standing on a test knowing the correlation between test and criterion

			-3σ		-2σ		-1σ		M		$+1\sigma$		$+2\sigma$		$+3\sigma$			
			20	25	30	35	40	45	50	55	60	65	70	75	80			
r	k	K														σ_{est}	r	
1.00	.00	0.0	20	25	30	35	40	45	50	55	60	65	70	75	80	.00	1.00	
.95	.31	3.1	21.5	26	31	36	40.5	45	50	55	59.5	64	69	74	78.5	.22	.95	
.90	.44	4.4	23	27.5	32	36.5	41	45.5	50	54.5	59	63.5	68	72.5	77	.30	.90	
.85	.53	5.3	24.5	29	33	37	41.5	46	50	54	58.5	63	67	71	75.5	.36	.85	
.80	.60	6.0	26	30	34	38	42	46	50	54	58	62	66	70	74	.40	.80	
.75	.66	6.6	27.5	31	35	39	42.5	46	50	54	57.5	61	65	69	72.5	.43	.75	
.70	.71	7.1	29	32.5	36	39.5	43	46.5	50	53.5	57	60.5	64	67.5	71	.46	.70	
.65	.76	7.6	30.5	34	37	40	43.5	47	50	53	56.5	60	63	66	69.5	.48	.65	
.60	.80	8.0	32	35	38	41	44	47	50	53	56	59	62	65	68	.49	.60	
.55	.84	8.4	33.5	36	39	42	44.5	47	50	53	55.5	58	61	64	66.5	.50	.55	
.50	.87	8.7	35	37.5	40	42.5	45	47.5	50	52.5	55	57.5	60	62.5	65	.50	.50	
.45	.89	8.9	36.5	39	41	43	45.5	48	50	52	54.5	57	59	61	63.5	.50	.45	
.40	.92	9.2	38	40	42	44	46	48	50	52	54	56	58	60	62	.49	.40	
.35	.94	9.4	39.5	41	43	45	46.5	48	50	52	53.5	55	57	59	60.5	.48	.35	
.30	.95	9.5	41	42.5	44	45.5	47	48.5	50	51.5	53	54.5	56	57.5	59	.46	.30	
.25	.97	9.7	42.5	44	45	46	47.5	49	50	51	52.5	54	55	56	57.5	.43	.25	
.20	.98	9.8	44	45	46	47	48	49	50	51	52	53	54	55	56	.40	.20	
.15	.99	9.9	45.5	46	47	48	48.5	49	50	51	51.5	52	53	54	54.5	.36	.15	
.10			47	47.5	48	48.5	49	49.5	50	50.5	51	51.5	52	52.5	53	.30	.10	
.05			48.5	49	49	49	49.5	50	50	50	50.5	51	51	51	51.5	.22	.05	
.00	1.00		50	50	50	50	50	50	50	50	50	50	50	50	50	.00	.00	

proximation. The zone of approximation is the zone in which the chances are 68 in 100 that a person with a given test score will be placed on the criterion. The column headed K gives the same data multiplied by 10. An example follows, illustrated in Figure 28.

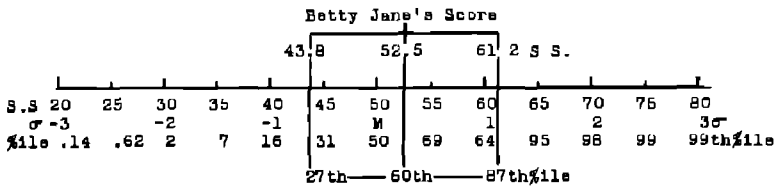


FIGURE 28

THE ZONE OF APPROXIMATION

Let us suppose that the correlation between the score on a test used in selecting packers and number of boxes packed per hour is .50. Applicant Betty Jane makes a standard score of 55 on the test (69th percentile) when compared with the criterion group of applicants for such work. Locating 55 on the scale at the top of Table 35 we follow it down the corresponding column to the row opposite .50 in the column headed 'r" (.50 because this is the validity coefficient for this test when used for this purpose). The figure at which we stop is 52.5. The standard error of estimate (k) corresponding to a validity coefficient of .50 is shown in column two to be .87 sigma of the number of boxes packed. As standard scores are used, the standard deviation of the scores is 10, and the standard

error of estimate in score units is 8 7 The most probable production score of Betty Jane is therefore 52 5 (60th percentile), and her zone of approximation is 52 5 ± 8 7 or 43 8 to 61 2 This is shown graphically in Figure 28 This figure brings out clearly the rough nature of the estimate provided by a validity coefficient of 50 standard scores of 44 and 61 correspond to percentiles of 27 and 87 In other words, there are 68 chances in 100 that Betty Jane's standing as a packer, if she is employed as such, will be somewhere between the 27th and 87th percentiles when workers are ranked according to output She may be a low average, an average, or a superior worker, although she is most likely to be high average And as there are only 68 chances in 100 that she will be found that effective there are also 16 chances in 100 that she will be less effective than that, placing somewhere in the bottom quarter of packers and 16 chances in 100 that she will prove better even than the 87th percentile placing near the top of the group in number of boxes packed per hour

To summarize these facts briefly, Betty Jane, who made a high average score on a test which is as valid for its purpose as most tests now in use, may prove after employment to be a low average average high average or superior worker The odds are 2 to 1 that she will be in one of these categories But she might turn out to be either one of the least effective workers or one of the best producers in the plant the odds are only 5 to 1 against either of these proving to be the case Such probabilities are useful as they give one a definite basis for making a decision but they clearly provide only a 'best estimate' of what Betty Jane will do not a prediction

APPENDIX B

TEST PUBLISHERS AND SCORING SERVICES REFERRED TO IN TEXT

American Council on Education
(see Educational Testing Service)

American Institute for Research
Cathedral of Learning
Pittsburgh 13, Pennsylvania

Association of American Medical Colleges (see Educational Testing Service)

Bureau of Educational Research and Service
University of Iowa
Iowa City, Iowa

C H Stoelting and Company
424 North Homan Avenue
Chicago, Illinois

California Test Bureau
5916 Hollywood Boulevard
Los Angeles 28, California

Center for Psychological Service
George Washington University
2026 G Street, N W
Washingtin, D C

Cooperative Test Service
15 Amsterdam Avenue
New York 23, New York

Division of Applied Psychology
Purdue University
Lafayette Indiana

Educational Records Bureau
437 West 59th Street
New York 24, New York

Educational Test Bureau
720 Washington Avenue, S E
Minneapolis, Minnesota

Educational Testing Service
Box 592
Princeton New Jersey

Engineers Northwest
100 Metropolitan Life Bldg
Minneapolis 1, Minn

Grune and Stratton
381 Fourth Avenue
New York, New York

Harvard University Press
Cambridge Massachusetts

Houghton Mifflin Company
2 Park St
Boston 7 Mass

McKnight and McKnight
Bloomington, Illinois

Marietta Apparatus Company
Marietta, Ohio

Psychological Corporation
522 Fifth Avenue
New York City New York

Psychological Institute
Lake Alfred
Florida

Public School Publishing Company
Bloomington, Illinois

Science Research Associates
228 South Wabash Avenue
Chicago Illinois

Sheridan Supply Company
Beverly Hills
California

Stanford University Press
Stanford University California

United States Air Force (Aviation Psychology Program)
Washington, D C

United States Employment Service
Washington D C

University of Iowa
Iowa City Iowa

University of Minnesota Press
Minneapolis Minnesota

West Publishing Company
St Paul Minnesota

Williams and Wilkins
Baltimore Maryland

World Book Company
Yonkers on Hudson New York

BIBLIOGRAPHY

1 ABEL, T M A study of a group of subnormal girls successfully adjusted in industry and the community *Amer J ment Def*, 1940, *45*, 66–72

2 ACHILLES, P S The consulting services of the Psychological Corporation *J consult Psychol*, 1946, *10*, 120–126

3 ACHILLES, P S, and SCHULTZ, R S Characteristics of life insurance sales men *Pers J*, 1934, *12*, 260–263

4 ADAMS, W M Prediction of scholastic success in colleges of law I The experimental edition of the Iowa Legal Aptitude Test II An investigation *Educ Psychol Measmt*, 1944, *3* 291–305, and *4*, 13–19

5 ADKINS, D C Test construction in public personnel administration *Educ Psychol Measmt*, 1944 *4*, 141–160

6 ADKINS, D C Construction and analysis of written tests for predicting job performance *Educ Psychol Measmt*, 1946 *6* 195–212

7 ADKINS D C et al *Construction and analysis of achievement tests* Washington D C Government Printing Office 1947

8 ADKINS, D C and KUDER, G F The relation between primary mental abilities and activity preferences *Psychometrika* 1940, *5*, 251–262

9 ALLEN A T Cogs in the occupational wheel *Occupations*, 1941, *20* 15–18

10 ALLEN, C M *Some effects produced in an individual by knowledge of his own intellectual level* New York Teachers College Columbia Univ, 1930

11 ALLEN L P, and SMITH P *The value of vocational tests as aids to choice of employment* Birmingham, Engl City of Birmingham Educ Committee, 1940

12 ALLPORT G W *Personality a psychological interpretation* New York Henry Holt, 1937

13 ALLPORT, G W The psychologist s frame of reference *Psychol Bull*, 1940, *37*, 16–17

14 ALTFNFDER, L E The value of intelligence, personality, and vocational interest tests in a guidance program *J educ Psychol*, 1940, *31*, 449–459

15 ALIUS, W D and MAHLER C \ Significance of verbal aptitude in the type of occupation pursued by illiterates *J appl Psychol*, 1946, *30*, 155–160

16 ANDERSON, E E, et al Wilson College studies in psychology I A comparison of the Wechsler Bellevue, Revised Stanford Binet, and American

Council on Education tests at the college level *J Psychol*, 1942, *14*, 317–326

17 ANDERSON, R G Technological aspects of counseling adult women *J appl Psychol*, 1938, *22*, 455–469

18 ANDERSON R G Abbreviated forms of the Bernreuter Personality Inventory *J consult Psychol* 1947, *11* 310–314

19 ANDERSON, R G Test scores and efficiency ratings of machinists *J appl Psychol*, 1947, *31*, 377–388

20 ANDERSON, V V *Psychiatry in industry* New York Harper & Brothers, 1929

21 ANDREW, D M An analysis of the Minnesota Vocational Test for Clerical Workers I II *J appl Psychol*, 1937, *21*, 18–47 139–172

22 ANDREW, D M, and PATERSON, D G Measured characteristics of clerical workers Univ of Minnesota *Bulletin of the Employment Stabilization Research Institute*, No 1, 1934

23 Anonymous The question of allowing two days for the examinations of the College Entrance Examination Board Princeton, N J College Entrance Examination Board Report (Mimeo), 1947

24 Anonymous The 19—— college sophomore testing program *Educ Rec* (annually, October, since 1932)

25 Anonymous Tables of norms and reliability coefficients New York Co operative Test Service, 1935

26 Anonymous Testing is big business *Amer Psychol*, 1947, 2 26

27 Anonymous Results of Orientation Test in schools of business of fifty colleges American Institute of Accountants, College Testing Program, Bulletin No 4 New York Committee on Selection of Personnel, 1947

28 Anonymous Tests and testing programs Princeton, N J Educational Testing Service, 1948

29 ARSLNIAN, S A further study of the Clecton Vocational Interest Inventory *Occupations*, 1941, *20* 94–99

30 ARSENIAN, S Own estimate and objective measurement *J educ Psychol*, 1942, *33*, 291–302

31 ARSENIAN, S Informing college freshmen of their test scores *Sch and Soc*, 1942, *55*, 396–399

32 ARSENIAN S Change in evaluative attitudes during four years of college *J appl Psychol*, 1943 *27*, 338–349

33 Assessment Staff *Assessment of men selection of personnel for the Office of Strategic Services* New York Rinehart, 1948

34 AWAJI, Y Intelligenzpruefung in Japanischen Heere *Zeitsch fur angew Psychol*, 1928, *30*, 81–118

35 BABCOCK, H, and EMERSON, M R An analytical study of the MacQuarrie Test of Mechanical Ability *J educ Psychol*, 1938, 29, 50–56

36 BAGGALEY, A R The relation between scores obtained by Harvard freshmen on the Kuder Preference Record and their fields of concentration *J educ Psychol*, 1947, *38*, 421–427

37 BAILEY, H W, et al Counseling and the use of tests in the Student Personnel Bureau at the University of Illinois *Educ Psychol Measmt*, 1946, *6*, 37–60

38 BALDWIN, E F, and SMITH, C F Performance of adult female applicants for factory work on the Revised Minnesota Paper Form Board Test *J appl Psychol*, 1944, *28* 468–470

39 BALINSKY B An analysis of the mental factors of various age groups from nine to sixty *Genet Psychol Monog*, 1941 *23*, 191–234

40 BALL, F J A study of the predictive values of the Thurstone Primary Mental Abilities as applied to lower division freshmen Unpubl thesis Penn State College 1940

41 BALL R S Predictability of occupational level from intelligence *J consult Psychol*, 1938 *2*, 184–186

42 BARNES, M W Relationship of the study of mathematics to Q scores on the A C E Psychological Examination *Sch Sci Math*, 1943 *43* 581–582

43 BARNES, M W Gains in the A C E Psychological Examination during the freshman sophomore years *Sch and Soc* 1943 *57*, 250–252

44 BARNETTE W L Norms of business college students on standardized tests Intelligence, clerical ability English *J appl Psychol* 1940 *24* 237–243

45 BARRETT D W Aptitude and interest patterns of art majors in a liberal arts college *J appl Psychol*, 1945 *29*, 483–492

46 BARRETT, D W Prediction of achievement in typewriting and stenography in a liberal arts college *J appl Psychol*, 1946 *30*, 624–630

47 BARRY, C W Kuder Preference Record norms based on measurements made on high school seniors *Occupations*, 1944 *22*, 487–488

48 BARTLETT, E D Aptitude test and the selection of office workers *Amer Mgt Assn, Off Mgt Series*, No 84 1938 4–12

49 BATES J WALLACE, M and HENDERSON M I A statistical study of four mechanical ability tests *Proc Ia Acad Sci*, 1943 *50* 299–301

50 BAUMAN M K Studies in the application of motor skills techniques to the vocational adjustment of the blind *J appl Psychol*, 1946, *30*, 144–154

51 BAUMGARTEN, F Behavior in taking tests *Occupations*, 1935 *15*, 115–122

52 BAXTER, B An experimental analysis of the contributions of speed and level in an intelligence test *J educ Psychol*, 1941, *32*, 285–296

53 BAXTER, B On the equivalence of time limit and work limit methods *Amer J Psychol*, 1942 *55*, 407–411

54 BEATTY, J D and CLETON G U Predicting achievement in college and after graduation *Pers J*, 1928, *6*, 344–351

55 BECK, S J Introduction to the Rorschach method *Amer Orthopsychiat Assn*, Monog No 1, 1937

56 BECK, S J *Rorschach's test I Basic processes* New York Grune and Stratton, 1944

57 BECK, S J *Rorschach's test II A variety of personality pictures* New York Grune and Stratton, 1945

58 BECKHAM A S Employment and guidance minimum intelligence levels for several occupations *Pers J*, 1930, *9* 309–313

59 BEDELL R The relationship between self-estimated and measured vocational interests *J appl Psychol*, 1941 *25*, 54–66

60 BEERY, G S An experiment in self analysis *J educ Psychol*, 1946 *37*, 111–124

61 BELL, H M *Youth tell their story* Washington, D C American Council on Education, 1938

62 BELL, H M *The theory and practice of personal counseling* Palo Alto Stanford Univ Press, 1935 1939

63 BELL, H M The Bell Adjustment Inventory (in Kaplan O J *Encyclopedia of vocational guidance*) New York Philosophical Library, 1948

64 BELLAK, L *Bellak TAT Blank and Manual* New York Psychological Corporation, 1947

65 BELLOWS, R M Status of selection and counseling techniques for dental students *J consult Psychol*, 1940 *4*, 10–14

66 BELLOWS R M Procedures for evaluating vocational criteria *J appl Psychol* 1941, *25* 499–513

67 BELLOWS R M and KAPLAN, M Interest patterns of good and poor dental students a new scoring key for the Strong Vocational Interest Blank *Psychol Bull* 1941, *38*, 608 (abstract)

68 BENNETT, G K, and CRUIKSHANK, R M *A summary of manual and mechanical ability tests* New York Psychological Corporation 1940

69 BENNETT G K, and CRUIKSHANK R M Sex differences in the understanding of mechanical problems *J appl Psychol*, 1942 *26* 121–127

70 BENNETT G K, and ILAR, R A Mechanical comprehension and dexterity *Pers J* 1943 *22* 12–17

71 BENNETT, G K, and GORDON, H P Personality test scores and success in nursing *J appl Psychol*, 1944, *20*, 267–278

72 BENNETT, G K, and WESMAN, A G Industrial test norms for a Southern plant population *J appl Psychol*, 1947, *31*, 241–246

73 BENTON, A L The Minnesota Multiphasic Inventory in clinical practice *J nerv Mental Dis*, 1945 *102*, 416–420

74 BINTON, A L, and PROBST, K A comparison of psychiatric ratings with Minnesota Multiphasic Personality Inventory scores *J abnorm soc Psychol*, 1946, *41*, 75–78

75 BENTON, A L, and PERRY, J D A study of the predictive value of the Stanford Scientific Aptitude Test *J Psychol* 1940, *10*, 309–312

76 BERDIE, R F Likes, dislikes and vocational interest *J appl Psychol*, 1943 *27*, 180–189

77 BERDIE, R F Factors associated with vocational interests *J educ Psychol*, 1943, *34*, 257–277

78 BERDIE, R F Prediction of college achievement and satisfaction *J appl Psychol*, 1944, *28*, 239–245

79 BERDIE, R F Factors related to vocational interests *Psychol Bull*, 1944 *41*, 137–157

80 BERDIE R F Range of interests and psychopathologies *J clin Psychol*, 1946, 2, 161–166

81 BERG I H A study of success and failure among student nurses *J appl Psychol*, 1947 *31*, 389–396

82 BERGEN G L SCHNEIDLER G and SHERMAN I *Use of tests at the Adjustment Service* New York Amer Assn for Adult Educ 1935

83 BERKSHIRE, J R, et al Test preferences in guidance centers *Occupations*, 1938 *26*, 337–343

84 BERMAN, I R, DARLEY J G, and PATERSON D G Vocational interest scales Univ of Minnesota, *Bull Empl Stab Res Inst* No 5, 1934

85 BERNARD J Some biological factors in personality and marriage *Human Biology* 1935 *7*, 430–436

86 BERNREUTER, R G The validity of the Personality Inventory *Pers J*, 1933, *11* 383–386

87 BERNREUTER, R G The theory and construction of the Personality Inventory *J soc Psychol* 1933 *4* 387–405

88 BERNREUTER R G and GOODMAN C H A study of the Thurstone Primary Mental Abilities Tests applied to freshman engineering students *J educ Psychol*, 1941, *32*, 55–60

89 BERNREUTER, R G and JACKSON, T A Sales personnel selection and related services *J consult Psychol*, 1946 *10* 127–140

90 BILLS, M A Relation of scores in Strong's Interest Blank to success in selling casualty insurance *J appl Psychol*, 1938, 22 97–104

91 BILLS, M A Selection of casualty and life insurance agents *J appl Psychol*, 1941, 25, 6–10

92 BILLS M A and WARD, L W Testing salesmen of casualty insurance *Pers J*, 1936 *15*, 55–58

93 BINGHAM, W V Classifying and testing for clerical jobs *Pers J*, 1935, *14*, 163–172

94 BINGHAM, W V *Aptitudes and aptitude testing* New York Harper & Brothers, 1937

95 BINGHAM, W V and DAVIS, W T Intelligence test scores and business success *J appl Psychol*, 1924 *8*, 1–22

96 BINGHAM W V, and MOORE, B V *How to interview* New York Harper & Brothers, 1931 1934 and 1941

97 BIXLER, R H, and BIXLER, V H Clinical counseling in vocational guidance *J clin Psychol*, 1945, *1*, 186–192

98 BIXLER, R H and BIXLER V H Test interpretation in vocational counseling *Educ Psychol Measmt* 1946, *6* 145–156

99 BLAIR, G M, and KAMMAM, F F Do intelligence tests requiring reading ability give spuriously low scores to poor readers at the college level? *J educ Res*, 1942, *36*, 280–283

101 BLOOM, B S Further validation of the Bernreuter Personality Inventory *Penn State Coll Studies in Education*, No 14, 1936

102 BLULTT, C H Vocational interests of vocational rehabilitation officers *Occupations*, 1945 *24* 25–32

103 BLUM M L A contribution to manual aptitude measurement in industry *J appl Psychol*, 1940 *24*, 381–416

104 BLUM M L Selection of sewing machine operators *J appl Psychol*, 1943, *27*, 35–40

105 BLUM, M L and CANDEE, B The selection of department store packers and wrippers with the aid of certain psychological tests *J appl Psychol*, *1941*, *25* 78–85

106 BLUM M L, and CANDEE B The selection of department store packers and wrappers with the aid of certain psychological tests Study II *J appl Psychol*, 1941, *25*, 291–299

107 BOARD E et al Relation of general intelligence to mechanical ability *Industrial Arts*, 1927, *16*, 330–342

108 BOCHNER R and HALPERIN F *The clinical application of the Rorschach Test* New York Grune and Stratton, 1945 (rev ed)

109 BOLANOVICH D J and GOODMAN, C H A study of the Kuder Preference Record *Educ Psychol Measmt*, 1944, *4* 315–325

110 BOLENBAUGH L and PROCTOR, W M Relation of the subjects taken in high school to success in college *J educ Res*, 1927, *15*, 87–92

111 BORDIN E S A theory of vocational interests as dynamic phenomena *Educ Psychol Measmt*, 1943, *3*, 49–66

112 BORDIN, F S Relative correspondence of professed interests to Kuder and Strong interest test scores *Amer Psychol*, 1947, 2 293 (abstract)

113 BORDIN, E S, and BIXLER R H Test selection a process of counseling *Educ Psychol Measmt* 1946 *6*, 361–374

114 BOROW, H Growth and present status of occupational testing *J consult Psychol*, 1944, *8*, 70–79

115 BRADSHAW, F F The American Council on Education Rating Scale *Arch Psychol*, No 119, 1930

116 BRAGDON, H D *Counselling the college student* Cambridge Harvard Univ Press, 1929

117 BRANSFORD T L, et al A study of the validity of written tests for administrative personnel *Amer Psychol*, 1946, 7 279 (abstract)

118 BREWER J M Religion and vocational success *Relig Educ* 1930 25, 1–8

119 BREWER, J M *Education as guidance* New York Macmillan 1932

120 BREWER, J M Classification of items in interest inventories *Occupations* 1943 21, 448–451

121 BROWN J F *The psychodynamics of abnormal behavior* New York McGraw Hill, 1940

122 BRUSH EDWARD N Mechanical ability as a factor in engineering aptitude *J appl Psychol*, 1941 25, 300–312

123 BRYAN A I Grades intelligence and personality of art school freshmen *J educ Psychol*, 1942 33, 50–64

124 BURGEMEISTER B B The permanence of interests of women college students *Arch Psychol* No 255, 1940

125 BURNHAM P S Stability of interests *Sch and Soc* 1942 55, 332–335

126 BUROS O K (Ed) *Mental measurement yearbook* New Brunswick, N J Rutgers Univ Press 1938 1939 1940 1949

127 BUXTON I Application of multiple factorial methods to the study of motor abilities *Psychometrika* 1938 3 85–93

128 BYRNS R Relation of vocational choice to mental ability and occupational ability *Sch Rev*, 1939 47, 101–109

129 BYRNS R and HENMON, V A C Parental occupation and mental ability *J educ Psychol*, 1936, 27 281–291

130 CAHOON A P Use of the Bernreuter Personality Inventory in the selection of student teachers *Univ High School J* 1933, 13 91–103

131 CAITIS R and WRENN C J The GED Tests as predictors of scholastic success *Educ Psychol Measmt*, 1947 7 93–100

132 CAMPBELL WM Measurement in determining the personality of the college cribber *Educ*, 1933 53, 403–408

133 CANDEE B, and BLUM M L Report of a study done in a watch factory *J appl Psychol*, 1937, 21, 572–582

134 CANDEE B and BLUM M L A new scoring system for the Minnesota Clerical Test *Psychol Bull*, 1937 34 545 (abstract)

135 CANNING L et al Permanence of vocational interests in high school boys *J educ Psychol*, 1941 32, 482

136 CANTRIL H and ALLPORT G W Recent applications of the Study of Values *J abnorm soc Psychol*, 1933, 28, 259–273

137 CAPWELL, DORA F Personality patterns of adolescent girls I, II *J appl Psychol*, 1945, *29*, 212–228 289–297

138 CARDALL, A J A test for primary business interests based on a functional occupational classification *Educ Psychol Measmt* 1942, *2*, 113–138

139 CARROLL, H A What do the Meier Seashore and the McAdory Art Tests measure? *J educ Res*, 1933 *26* 661–665

140 CARROLL, H A *Genius in the making* New York McGraw Hill, 1940

141 CARROLL, H A, and ZURICH A C Intelligence and art appreciation *J educ Psychol*, 1932 *23* 214–221

142 CARTER, H D Twin similarities in occupational interests *J educ Psychol*, 1932, *23* 641–655

143 CARTER, H D Twin similarities in emotional traits *Char Pers*, 1935 *4*, 61–78

144 CARTER, H D The development of vocational attitudes *J consult Psychol*, 1940 *4*, 185–191

145 CARTER, H D Vocational interests and job orientation Palo Alto Stanford Univ Press *Applied Psychol Monog*, No 2 1944

146 CARTER H D The development of interest in vocations *Yrbk, nat Soc Stud Educ* 1941, *43* 255–276

147 CARTER H D, and JONES M C Vocational attitude patterns in high school students *J educ Psychol* 1938 *29* 321–334

148 CARTER H D and STRONG E K JR Sex differences in the occupational interests of high school students *Pers J* 1933 *12* 166–175

149 CARTER H D TAYLOR K V F and CANNING I B Vocational choices and interest test scores of high school students *J Psychol*, 1941, *11*, 297–306

150 CATTELL PSYCHE Why Otis "IQ" cannot be equivalent to the Stanford Binet I Q *J educ Psychol*, 1931 *22*, 509–603

151 CATTELL R B Occupational norms of intelligence *Brit J Psychol*, 1934, *25* 1–28

152 CAVLTI, J W, et al Tests of medical aptitude at Minnesota *J Assn Amer Med Coll*, 1937, *12* 257–268

153 CHANNING A *Employability of mentally deficient boys and girls* Washington, D C Government Printing Office 1932

154 CHAPMAN, J C *Trade tests* New York Henry Holt 1921

155 CHESNEY, A M Evaluation of the Medical Aptitude Test *J Assn Amer Med Coll*, 1936 *2* 15–32

156 CHRISTENSEN A M Traits of college going, employed and unemployed high school graduates *Sch Rev* 1938, *46*, 597–601

157 CHRISTENSEN, T E Some observations with respect to the Kuder Preference Record *J educ Res*, 1946 *40* 96–107

158 CHRISTENSEN, T E Dictionary classification of AGCT scores for selected civilian occupations *Occupations*, 1946, *25*, 97–101

159 CLARK, C D, and GIST, N P Intelligence as a factor in occupational choice *Amer Sociol Rev*, 1938, *3*, 683–694

160 CLARK, W A, and SMITH, I F Further evidence on the validity of personality inventories *J educ Psychol*, 1942, *33* 81–91

161 CLEETON, G U Occupational adjustment in Allegheny County Pittsburgh *Bulletin of the Pittsburgh Personnel Assn*, 1935

162 COHEN, L, and STRAUSS, L Time study and the fundamental nature of manual skill *J consult Psychol*, 1946 *10*, 146–154

163 COINER, GLADYS E The personality of the college cribber Master's thesis Univ of S Cal, 1938

164 COLE, R C Evaluating a boys' club guidance program *Occupations*, 1939 *17*, 694–698

165 COLLINS, C J Development of a scoring key on the Strong Vocational Interest Inventory for women teachers of physical education *Res Quart Amer Assn Hlth phys Educ* 1942, *13*, 156–165

166 COMBS A W Nondirective techniques in vocational counseling *Occupations* 1947, 25, 262–267

167 COMPTON, R U Student evaluation of knowing college aptitude test score *J educ Psychol* 1941, 32 656–664

168 CONGDON N A A study of Cleeton's Vocational Interest Inventory *Occupations*, 1940, *18*, 347–352

169 CONRAD H S Summary report on research and development of the Navy's aptitude testing program Final report on contract OEM s r 705 OSRD 1945, Publ Bd No 13289 Washington D C Dept of Commerce 1946

170 COOK D W, and BARRI, W F Effect of specialized industrial norms on the use of the Minnesota Rate of Manipulation Test as a selective instrument *J appl Psychol*, 1942 *26* 785–792

171 COPELAND, H A Some characteristics of three tests used to predict clerical success *J appl Psychol* 1936 *20*, 461–470

172 CORSINI, R Bernreuter patterns of a group of prison inmates *J clin Psychol* 1946 2 284–285

173 COYNER B J Nondirective interviewing techniques in vocational counseling *J consult Psychol* 1947 *11* 70–74

174 COWDERY, K M Measures of professional attitudes *J Pers Res*, 1926 *5*, 131–141

175 COX, J R *Mechanical aptitude* London Methuen 1928

176 COX, J R *Manual aptitude* London Methuen 1934

177 COXE, W W, and CORNELL, E L Prognosis of teaching ability of students in New York State normal schools *Univ of State of N Y Bull*, No 1033 Albany, 1934

178 CRAWFORD, A B Some observations on primary mental abilities battery in action *Sch and Soc*, 1940, *51*, 585–592

179 CRAWFORD, A B, and BURNHAM P S Trial at Yale University of the USAFI General Educational Development Tests *Educ Psychol Measmt*, 1944, *4* 261–270

180 CRAWFORD A B and BURNHAM P S *Forecasting college achievement a survey of aptitude tests for higher education* Part I General considerations in the measurement of academic promise New Haven Yale Univ Press, 1946

181 CRIDER B A school of nursing selection program *J appl Psychol*, 1943, *27*, 452–457

182 CRISSY, W J E and DANIEL, W J Vocational interest factors in women *J appl Psychol*, 1939 *23*, 488–494

183 CROOKS W R and FERGUSON L W Item validities of Otis Self Administering Tests of Mental Ability for colleges *J exp Educ*, 1941 *9*, 229–232

184 CROSBY R C Scholastic achievement and measured interests *J appl Psychol*, 1943 *27* 101–103

185 CROSBY R C and WINSOR A L The validity of student estimates of their interests *J appl Psychol* 1941 *25* 408–414

186 CURETON, E E Evaluation or guidance? *J exp Educ*, 1940, *8*, 308–340

187 DARLEY J G Reliability of tests in the standard battery (In Research studies in individual diagnosis) Univ of Minnesota *Bull Empl Stab Res Inst*, No 4 1934

188 DARLEY J G Preliminary study of the relationships between attitudes, adjustment and vocational interest tests *J educ Psychol*, 1938 *29*, 467–473

189 DARLEY J G *Clinical aspects and interpretation of the Strong vocational interest blank* New York Psychological Corporation 1941

190 DARLEY J G *Testing and counseling in the high school guidance program* Chicago Science Research Associates 1943

191 DARLEY J G Relationships among the Primary Mental Abilities Tests selected achievement measures personality tests and tests of vocational interests University of Minnesota, *Studies in Higher Education*, 1941, 192–200

192 DARLEY, J G and MCNAMARA W J Factor analysis in the establishment of new personality tests *J educ Psychol*, 1940, *31*, 321–334

193 DAVIDSON C M Evaluation of clerical tests *Pers J*, 1937, *16*, 57–64, 95–98

194 DAVIDSON W M and CARROLL J B Speed and level components in time limit scores *Educ Psychol Measmt*, 1945, *5*, 411–428

195 DAVIS, F B (Ed) The AAF qualifying examination *AAF Aviation Psychology Report*, No 6 Washington Government Printing Office, 1947

196 DAVIS, F B *Item analysis* Cambridge Harvard Univ Press, 1947

197 DEARBORN, W F, and ROTHNEY J W M Scholastic economic and social backgrounds of unemployed youth *Harvard Bull in Educ*, No 20, 1938

198 DERFLINGER, G W The prediction of college success—a summary of recent findings *J Amer Assn coll Registrars* 1947 *19* 68–78

199 DETCHEN L The effect of a measure of interest factors on prediction of performance in a college social sciences comprehension examination *J educ Psychol*, 1946 *37*, 45–52

199a DIAMOND, S The interpretation of test profiles *J appl Psychol* 1948 *32*, 512–520

200 DIMICHAEL S G and DABELSTEIN D H Work satisfaction and efficiency of vocational rehabilitation counselors as related to measured interests *Amer Psychol*, 1947 *2* 342 (abstract)

201 DODGE A F Occupational ability patterns New York *Teachers College Contributions to Education* No 658 1935

202 DODGE, A F Social dominance of clerical workers and salespersons as measured by the Bernreuter Personality Inventory *J educ Psychol*, 1937, *20*, 71–73

203 DODGE A F What are the personality traits of the successful clerical worker? *J appl Psychol* 1940 *24*, 576–586

204 DOUGLASS H R Different levels and patterns of ability necessary for success in college *Occupations* 1943 *22* 182–186

205 DOUGLASS H R Prediction of success in the Medical School *Univ Minn Stud Predict schol Ach* 1942 *2* 1–16

206 DOUGLASS H R et al Prediction of success in the Law School *Univ Minn Stud Predict schol Ach* 1942 *2* 46–60

207 DOUGLASS H R and MAASKE R S Prediction of success in the School of Business Administration *Univ Minn Stud Predict schol Ach*, 1942, *2* 32–45

208 DOUGLASS H R, and MCCULLOUGH C M Prediction of success in the School of Dentistry *Univ Minn Stud Predict schol Ach*, 1942 *2*, 61–74

209 DOUGLASS H R, and MERRILL R G Prediction of success in the School of Nursing *Univ Minn Stud Predict schol Ach* 1942 *2* 17–31

210 DRAKE R M Validity and reliability of tests of musical talent *J appl Psychol*, 1937, *17* 447–458

211 DRAKE R M Factor analysis of music tests *Psychol Bull*, 1953 *36*, 608–609

212 DREPS H F Psychological capacities and abilities of college art students of high and low standing *Psychol Monog* 1933 *45* 134–146

213 DROUGHT N E An analysis of eight measures of personality adjustment in relation to relative scholastic achievement *J appl Psychol*, 1938, *22*, 597–606

214 DuBois P H (Ed) The classification program *AAF Aviation Psychology Report*, No 2 Washington, D C Government Printing Office, 1947

215 Dudycha C J Dependability and clerical aptitude *J appl Psychol*, 1939, 23 332 336

216 Duffy Elizabeth A critical review of investigations employing the Allport Vernon Study of Values and other tests of evaluative attitude *Psychol Bull* 1940 37 597–612

217 Duffy E and Crissy, W J E Evaluative attitudes as related to vocational interest and academic achievement *J abnorm soc Psychol*, 1940, *35*, 226–245

218 Dulsky, S G Vocational counseling I By use of tests *Pers J*, 1941, 20, 16–22

219 Dunlap J W Preferences as indicators of specific academic achievement *J educ Psychol*, 1935 *6* 411–415

220 Dunlap J W The predictive value of interest test items for achievement in various school subjects *J appl Psychol* 1935, *19*, 53–58

221 Dunlap J W and Harper B P Profiles of interest scores *J higher Educ*, 1944 *15* 159–160

222 Dunlop F S Subsequent careers of non academic boys New York *Teachers College Contributions to Education* 1935

223 Dvorak B J Differential occupational ability patterns Univ of Minnesota *Bull Empl Stab Res Inst* No 8 1935

224 Dvorak B J The new U S E S General Aptitude Test Battery *J appl Psychol* 1947 *31* 372–376

225 Dvorak B The U S E S General Aptitude Test Battery *Occupations*, 1947 *26*, 42–44

226 Dyer D T Relation between vocational interests of men in college and their subsequent occupational histories for ten years *J appl Psychol*, 1939 23 280–288

227 Dyer H S Evidence on the validity of the USAFI Tests of General Educational Development *Educ Psychol Measmt* 1945 5 321–333

228 Dysinger D W A critique of the Humm Wadsworth Temperament Scale *J abnorm soc Psychol* 1939, *34*, 78–85

229 Dysinger W S Test reports in educational counseling *Educ Psychol Measmt*, 1943 3 361–366

230 Dysinger W S Two vocational diagnoses compared *Occupations* 1944, 22, 304–308

231 Dysinger, W S Use of tests at McMurray College *Educ Psychol Measmt*, 1946 *6*, 61–72

232 Earle F M *Psychology and the choice of a career* London Methuen, 1933

233 Ebert E and Simmons K The Brush Foundation study of child growth

and development I Psychometric tests *Monog Soc Res Child Development*, No 2, 1943

234 ECKERT, R, and MARSHALL, T O *When youth leave school* New York McGraw-Hill, 1938

235 EDGERTON, H A, and BRITT, S H The science talent search *Occupations*, 1943, 22, 177–180

236 EDLUND S W, and EDLUND, M G *Pick your job and land it* New York Prentice Hall, 1938

237 ELLS, W C Scholastic ability of secondary school pupils *Educ Rec*, 1937 *18*, 58–67

238 ELLIS, A The validity of personality questionnaires *Psychol Bull* 1946, *43*, 385–440

239 ELLISON M L, and EDGERTON H A The Thurstone Primary Mental Abilities Tests and college marks *Educ Psychol Measmt*, 1941 *1*, 399–406

240 ESTES, S G A study of five tests of "spatial ability" *J Psychol*, 1942, *13* 265–271

241 ESTES, S G, and HORN D Interest patterns as related to fields of concentration among engineering students *J Psychol* 1939, *7* 29–36

242 EURICH A C and CARROLL, H A Group differences in art judgment *Sch and Soc* 1931 *34* 201

243 EVANS M C Social adjustment and interest scores of introverts and extroverts *Educ Psychol Measmt*, 1947 *7* 157–167

244 FAIRBANKS, R E The subnormal child—seventeen years after *Ment Hyg* 1933, *17*, 177–208

245 FARAGO, L (Ed) *German psychological warfare* New York Committee for National Morale 1941

246 FARNSWORTH P R An historical critical and experimental study of the Seashore Kwalwasser test battery *Genet Psychol Monog*, 1931, *9* 291–389

247 FARNSWORTH P R, and FERGUSON L W The growth of a suicidal tendency as indicated by score changes in Bernreuter's Personality Inventory *Sociometry*, 1938, *1* 330–341

248 FARNSWORTH, P R and MISUMI ISSEI Notes on the Meier-Seashore Art Judgment Test *J appl Psychol*, 1931 *15* 418–420

249 FAUBION, R W, CLEVELAND E A, and HARRELL W Influence of training on mechanical aptitude test scores *Educ Psychol Measmt*, 1942, *2*, 91–94

250 FAY, C J and MIDDLETON W I Relationship between musical talent and preferences for different types of music *J educ Psychol*, 1941, *32*, 573–589

251 FEDER, D D and BAER, L O A comparison of test records and clinical

evaluations of personal adjustment *J educ Psychol*, 1941, *32*, 133–144

252 FEDER, D D and MALLETT, D R The validity of certain measures of personality adjustment *J Amer Assn Reg*, 1937, *13*, 5–15

253 FERGUSON, L W, HUMPHREYS, L G, and STRONG, F W A factorial analysis of interests and values *J educ Psychol*, 1941, *32*, 197–204

254 FERSON, M L, and STODDARD, G D Law Aptitude Examination—preliminary report *Am Law Sch Rev*, 1927, *6*, 78–81

255 FINDLEY, W J A group testing program for the modern school *Educ Psychol Measmt*, 1945 *5*, 173–179

256 FISCHER, R P Emotional maladjustment and academic personnel *J higher Educ*, 1944, *15*, 43–44

257 FISHER, V E, and HANNA, J V *The dissatisfied worker* New York Macmillan, 1931

258 FISHER, W, and HAYES, S P, JR Maladjustment in college, predicted by Bernreuter Inventory scores and family position *J appl Psychol* 1941, *25* 86–96

259 FISKE, D W Naval aviation psychology IV The central research group *Amer Psychol* 1947, *2* 67–72

260 FISKE, D W Validation of Naval Aviation Cadet selection tests against training criteria *J appl Psychol*, 1947 *31*, 601–614

261 FLANAGAN, J C *Factor analysis in the study of personality* Palo Alto Stanford Univ Press, 1935

262 FLANAGAN, J C Measuring interest *Psychol Bull*, 1939, *36* 529–530

263 FLANAGAN, J C A preliminary study of the 1940 edition of the National Teachers Examination *Sch and Soc*, 1941 *54* 59–64

264 FLANAGAN, J C The experimental evaluation of a selection procedure *Educ Psychol Measmt*, 1946 *6* 445–466

265 FLANAGAN J C The Aviation Psychology Program in the AAF *AAF Aviation Psychology Report*, No 1 Washington, D C Government Printing Office, 1947

266 FLEMMING, E J and FLEMING C W A qualitative approach to the problem of improving selection of salesmen by psychological test *J Psychol*, 1946 *21*, 127–150

267 FLORY, C D, and JANNEY, J E Psychological services to business leaders *J consult Psychol*, 1946 *10*, 115–119

268 FORLANO, G and KIRKPATRICK F H Intelligence and adjustment measurements in the selection of radio tube mounters *J appl Psychol*, 1945, *29*, 257–261

269 FORRESTER, G Occupational pamphlets New York H W Wilson Co, 1948

270 FOSBERG, I A An experimental study of the reliability of the Rorschach Psychodiagnostic technique *Rorschach Res Exch*, 1941, *5*, 72–84

271 FRANDSEN, A Interests and general educational development *J appl Psychol*, 1947 *31*, 57–66

272 FRANDSEN A Appraisal of interests in guidance *J educ Res* 1945 *39*, 1–12

273 FREDERICKSEN N *Validity of Navy aptitude tests in service schools at the Great Lakes Naval Training Station* Washington D C Dept of Commerce 1946

274 FRIEND J G and HAGGARD E A Work adjustment in relation to family background Palo Alto Stanford Univ Press *Appl Psychol Monog* No 16 1948

275 FROEHLICH C P and BENSON A L *Guidance testing* Chicago Science Res Associates 1948

276 FRYER D Occupational intelligence standards *Sch and Soc* 1922 *16*, 273–277

277 FRYER D *Measurement of interests* New York Henry Holt 1931

278 FRYER D and SPARLING E J Intelligence and occupational adjustment *Occupations* 1934 *12* 55–63

279 GARRETSON O K Relationships between expressed preferences and curricular abilities of ninth grade boys New York *Teachers College Contributions to Education* No 396 1930

280 GARRETT H L Forecasting job efficiency *Pers J* 1942 *20* 276–278

281 GARRETT H E A developmental theory of intelligence *Amer Psychol* 1946 *1* 372 378

282 GARRETT H L *Statistics in psychology and education* New York Longmans Green 1947

283 GAUDET J J and MARRYOTT F S Predictive value of the Iverson Stoddard Law Aptitude Examination *Am Law Sch Rev* 1930 *7* 27–42

284 GERBERICH J R Validation of a state wide educational guidance program for high school seniors *Sch and Soc* 1931 *33* 606–610

285 GHISELLI E E Comparison of the Minnesota Vocational Test for Clerical Workers with the general clerical battery of the U S E S *J appl Psychol*, 1942 *26* 75–80

286 GHISELLI E L Tests for the selection of inspectors packers *J appl Psychol* 1942 *26* 468–476

287 GHISELLI L E Use of the Strong Vocational Interest Blank and the Pressey Senior Classification Test in the selection of casualty insurance salesmen *J appl Psychol* 1942 *26* 794–799

288 GHISELLI E L Use of the Minnesota Rate of Manipulation and O Connor Finger Dexterity Tests in the selection of package wrappers *J appl Psychol* 1943 *27* 33–44

289 GILBRETH F B and GILBRETH L M *Applied motion study* New York Sturgis and Walton 1917

290 GILGER, G A, JR Declaration of vocational interest *Occupations*, 1942, *20*, 276–279

291 GILKINSON, H and KNOWER F H A study of standardized personality tests and skill in speech *J educ Psychol*, 1941, *32*, 161–175

292 GLICK, H N Hurricane intelligence *Science*, 1940, *91*, 450

293 GOLDEN A L Personality traits of drama school students *Quart J Speech*, 1940, *26* 564–575

294 GOODENOUGH, F L New evidence on environmental influence on intelligence *Yrbk nat Soc Stud Educ* 1940, *39*, 307–365

295 GOODFELLOW, L D A study of the interests and personality traits of prospective teachers *Educ Admin Sup*, 1932 *18* 649–658

296 GOODMAN, C H Comparison of the interests and personality traits of engineers and liberal arts students *J appl Psychol* 1942 *26*, 721–737

297 GOODMAN, C H Prediction of college success by means of Thurstone's Primary Mental Abilities Test *Educ Psychol Measmt*, 1944, *4*, 125–140

298 GOODMAN C H The MacQuarrie Test for Mechanical Ability I Selecting radio assembly operators *J appl Psychol* 1916, *30*, 586–595

299 GOODMAN C H The MacQuarrie Test for Mechanical Ability II Factor analysis *J appl Psychol* 1917 *31* 150–151

300 GOODMAN C H The MacQuarrie Test for Mechanical Ability III Follow up study *J appl Psychol* 1917, *31* 502–510

301 GORDON H C, and HERRNESS W W Do vocational interest questionnaires yield consistent results? *Occupations*, 1942 *20* 424–429

302 GOTHAM A L Personality and teaching efficiency *J exp Educ*, 1945, *14*, 157–165

303 GOTTSDANKER R M Measures of potentiality for machine calculation *J appl Psychol*, 1943 *27*, 233–248

304 GOUGH H G Diagnostic patterns on the Minnesota Multiphasic Personality Inventory *J clin Psychol*, 1946, *2* 23–37

305 GRACE, A G The relationship of mental ability to occupational choices of adults *Voc Guid Mag* 1931, *10* 454–458

306 GREEN H J et al A manual of selected occupational tests for use in public employment offices University of Minnesota *Bull Empl Stab Res Inst* No 3 1933

307 GREENE E B A sampling of vocabularies of superior adults *J higher Educ*, 1938, *9* 484–489

308 GRELNE L B Vocabulary profiles of groups in training *J educ Res*, 1940, *33*, 569–575

309 GRFENE, E B *Measurements of human behavior* New York Odyssey Press, 1941

310 GREENE H A and JORGENSEN A N *Measurement and evaluation in the secondary school* New York Longmans Green, 1943

311 GREGORY, W S From high school to college *Occupations*, 1939, *18*, 190–198

312 GREGORY, W S Data regarding the Academic Interest Inventory *Educ Psychol Measmt*, 1946 *6* 475–490

313 GRIFFITHS, J R The relationship between scholastic achievement and personality adjustment of men college students *J appl Psychol* 1945 *29*, 360–367

314 GRIFFIN, C H and BOROW, H An engineering and physical science aptitude test *J appl Psychol*, 1944, *28* 376–387

315 GUILFORD, J P New standards for test evaluation *Educ Psychol Measmt*, 1946 *6*, 427–438

316 GUILFORD J P (Ed) Printed classification tests *AAF Aviation Psychology Report*, No 5 Washington D C Government Printing Office 1947

317 GUILFORD J P Factor analysis in a test development program *Psychol Rev*, 1948, *55*, 79–91

318 GUILFORD J P The Guilford personality inventories (In Kaplan O J *Encyclopedia of vocational guidance*) New York Philosophical Library 1948

319 GUILFORD J P, and ZIMMERMAN W S Some AAF findings concerning aptitude factors *Occupations*, 1947 *26*, 151–159

320 GUILFORD, J P and ZIMMERMAN W S The Guilford Zimmerman Aptitude Survey *J appl Psychol*, 1948 *32*, 24–34

321 HAHN, M E Notes on the Kuder Preference Record *Occupations*, 1945 *23* 467–470

322 HAHN, M L and KENDALL, W L Some comments in defense of non non directive counseling *J consult Psychol* 1947 *11* 74–81

323 HAHN M E and WILLIAMS C T The measured interests of Marine Corps Women Reservists *J appl Psychol* 1945 *29* 198–211

324 HALL O M Attitudes and unemployment *Arch Psychol* No 165 1934

325 HALL O M An aid to the selection of pressman apprentices *Pers J* 1940 *8* 77–81

326 HAMMOND, H P, and STODDARD G D A study of placement examinations *Univ of Iowa Studies in Educ*, No 7 1928

327 HAMPTON, P A comparative study of certain personality traits and success in retail selling *J appl Psychol* 1941, *25* 431–446

328 HANAWALT N G et al Leadership as related to Bernreuter personality measures II An item analysis of responses of college leaders and non-leaders *J soc Psychol*, 1944, *17*, 251–267

329 HANMAN B Performance of adult males on the Minnesota Paper Form Board Test and the O Rourke Mechanical Aptitude Test *J appl Psychol*, 1942, *26*, 809–811

330 HARDTKE, L V Aptitude testing for metal working occupations *Psychol Bull*, 1945, *42*, 679–694

331 HARGAN, J The reaction of native white convicts to the Bernreuter Personality Inventory *Psychol Clinic*, 1934, *22*, 138–140

332 HARMON, I R Test patterns in the vocational clinic *Educ Psychol Measmt* 1945 *7* 207–220

333 HARMON L R, and WITNER, D M Use of the Minnesota Multiphasic Personality Inventory in vocational adjustment *J appl Psychol*, 1945, *29*, 132–141

334 HARRELL, T W Validity of certain mechanical ability tests for selecting cotton mill machine fixers *J soc Psychol* 1947 *18*, 279–282

335 HARRELL T W A factor analysis of mechanical ability tests *Psychol Bull*, 1939 *36*, 524 abstract)

336 HARRELL T W A factor analysis of mechanical ability tests *Psychometrika*, 1940, *5*, 17–33

337 HARRELL T W AGCT results for Air Forces specialists *Educ Psychol Measmt*, 1946 *6*, 341–350

338 HARRELL, T W and FAUBION R Primary mental abilities and aviation maintenance courses *Educ Psychol Measmt*, 1941 *1* 59–66

339 HARRELL T W and FAUBION R Selection tests for aviation mechanics *J consult Psychol* 1940 *4* 104–105

340 HARRELL, J W, and HARRELL M S AGCT scores for civilian occupations *Educ Psychol Measmt* 1945 *5* 229–239

341 HARRIS A J Relative significance of measures of mechanical aptitude intelligence, and previous scholarship for predicting achievement in dental school *J appl Psychol* 1937, *21*, 513–521

342 HARRIS D Group differences in values within a university *J abnorm soc Psychol*, 1934 *29* 95–102

343 HARRIS R F and CHRISTIANSEN CAROLE Prediction of response to brief psychotherapy *J Psychol* 1946 *21* 269–284

344 HARRISON R Studies in the use and validity of the TAT with mentally disordered patients II III *Char Pers*, 1940 *9*, 122–138

345 HARRISON, R, and ROTTER J B Note on the reliability of the Thematic Apperception Test *J abnorm soc Psychol*, 1945, *50* 97–99

346 HARROWER M R and STEINER M E *Large scale Rorschach techniques* Springfield C C Thomas, 1945

347 HARTMANN G W Measuring teaching efficiency among college instructors *Arch Psychol*, No 154, 1933

348 HARTSON, L D Influence of level of motivation on the validity of intelligence tests *Educ Psychol Measmt*, 1945 *5* 273–283

349 HARTSON, L D, and SPROW, A J Value of intelligence quotients obtained

in secondary school for predicting college scholarship *Educ Psychol Measmt*, 1941, *1*, 387–398

350 HARVEY, O L Mechanical "aptitude" or mechanical ability? *J educ Psychol*, 1931, *22*, 517–522

351 HATHAWAY, S R The personality inventory as an aid in the diagnosis of psychopathic inferiors *J consult Psychol* 1939 *3*, 112–117

352 HATHAWAY, S R The Minnesota Multiphasic Personality Inventory (In Kaplan, O J *Encyclopedia of Vocational Guidance*) New York Philosophical Library, 1948

353 HATHAWAY, S R, and MCKINLEY, J C A multiphasic personality schedule I Construction of the schedule *J Psychol*, 1940, *10*, 249–254

354 HATHAWAY, S R and MCKINLEY, J C A multiphasic personality inventory III The measurement of symptomatic depression *J Psychol*, 1942, *14*, 73–84

355 HATHAWAY, S R and MCKINLEY, J C The Minnesota Multiphasic Personality Inventory V Hysteria hypomania, and psychopathic deviate *J appl Psychol* 1944, *28* 153–171

356 HATHAWAY, S R and MCKINLEY J C *Manual for the Minnesota Multiphasic Personality Inventory* New York Psychological Corporation, 1945

357 HAY E N Tests in industry II Practical illustration *Pers J*, 1941, *20*, 10–15

358 HAY E N Predicting success in machine bookkeeping *J appl Psychol* 1943, *27*, 483–493

359 HAY E N and BLAKEMORE, A M Testing clerical applicants *J appl Psychol*, 1942 *26*, 852–855

360 HAY L N, and BLAKLMORE A M The relationship between clerical experience and scores on the Minnesota Vocational Test for Clerical Workers *J appl Psychol*, 1944 *27* 311–315

361 HAYES, E G Selecting women for shop work, Hawthorne Works, Western Electric Co *Pers J*, 1932 *11* 69–85

362 HAYES, S P *Vocational aptitude tests for the blind* Watertown, Mass Perkins Institution, 1946

363 HAZELHURST, J H A factorial analysis of measures of mechanical aptitude *Psychol Bull* 1940, *37* 578 (abstract)

364 HELD O C A comparative study of the performance of Jewish and Gentile college students on the American Council Psychological Examination *J soc Psychol*, 1941 *13*, 407–411

365 HEMSATH, M E Theory and practice of temperament testing *Pers J*, 1939, *18*, 3–12

366 HERMANN M Predictive value of the Medical Scholastic Aptitude Test at Woman's Medical College *J Assn Amer Med Coll*, 1943, *18*, 113–117

367 HERTZ M R The Rorschach Inkblot Test historical summary *Psychol Bull*, 1935 32 33–66

368 HERTZ M R The validity of the Rorschach method *Amer J Orthopsychi*, 1941 11 512 519

369 HESSEMER M The Thurstone PMA Tests in a study of academic success with school of engineering chemistry and physics Unpubl thesis Penn State College 1942

370 HICHSMITH J A Selecting musical talent *J appl Psychol* 1929 13 486–493

371 HILDRETH G Occupational status and intelligence *Pers J* 1934 13, 158–157

372 HILDRETH G Comparison of early Binet records with college aptitude test scores *J educ Psychol* 1940 31 465–471

373 HILKERT R N Student counselors or hints from experience *Educ Rec* 1940 21 172–181

374 HINES M and O CONNOR J A measure of finger dexterity *J Pers Res* 1926 4 379–382

375 HOLCOMB G W and LAFFITT H R A prognostic study of engineering aptitude *J appl Psychol* 1942 16 107–115

376 HOPLOCK R Occupational ability patterns *Occupations* 1934 12 47–48

377 HOPLOCK R *Job satisfaction* New York Harper & Brothers 1935

378 HOPLOCK R (Ed) The criteria of vocational success—a symposium *Occupations* 1936 14 917–975

379 HORNE B M and McCALL W C A study of some local factors affecting students scores on the Minnesota Personality Scale *Educ Psychol Measmt* 1942 2 257 266

380 HORNING S D Testing mechanical abilities by the MacQuarrie Test *Ind Art Mag* 1926 25 348 350

381 HORSCH A C and DAVIS R A Personality traits of juvenile delinquents and adult criminals *J soc Psychol* 1938 9 57–66

382 HORST P Increasing the efficiency of selection tests *Pers J* 1931 12 254–259

383 HORST P *The prediction of personal adjustment* New York Social Science Res Council 1941

384 HOVLAND C I and WONDERLIC E F A critical analysis of the Otis Self Administering Test of Mental Ability *J appl Psychol* 1949 23 367–387

385 HULL C L *Aptitude testing* Yonkers N Y World Book Co 1928

386 HUMM D G Test validation on remote criteria *J appl Psychol* 1946 30 333–339

387 HUMM D G et al Combination scores for the Humm Wadsworth Temperament Scale *J Psychol* 1949 7 227–254

388 HUMM D G and WADSWORTH G W JR The Humm Wadsworth Temperament Scale *Amer J Psychiatry* 1935 *92* 163–200

389 HUNT, E G and SMITH G Vocational psychology and choice of employment *Occup Psychol* 1945 *19* 109–116

390 HUNT H C Why people lose their jobs or aren't promoted *Pers J* 1936 *15*, 227

391 HUNT J McV (Ed) *Personality and the behavior disorders* New York Ronald Press 1944

392 HUNT T Measuring teaching aptitude *Educ Admin Sup* 1929 *15* 334–342

393 HUNTER F C Changes in scores of college students on the A C E Psychological Examination at yearly intervals *J educ Res* 1942 *36* 281–291

394 HUNTER L C and JORDAN A M An analysis of qualities associated with leadership among college students *J educ Psychol* 1939 *30* 497–509

395 JACK L M An experimental study of ascendant behavior in preschool children *Univ of Iowa Studies in Child Welfare* No 3 1934

396 JACOBSEN T T An evaluation of certain tests in predicting mechanic learner achievement *Educ Psychol Measmt* 1943 *3* 259–267

397 JACOBSON C I Interest patterns and achievement in medical school *J Assn Amer Med Coll* 1942 *17* 153–164

398 JARVIE I I and ELLINGSON M *A handbook of the anecdotal behavior journal* Chicago Univ Chicago Press 1940

399 JENKINS J G Naval aviation psychology II The procurement and selection organization *Amer Psychol* 1946 *1* 45 49

400 JENKINS J G Validity for what *J consult Psychol* 1946 *10* 93 98

401 JENNINGS J R and STOTT M B A fourth follow up of vocationally advised cases *Human Factor* 1936 *10* 165–174

402 JOHANNSEN D F The effect of neuroticism upon the capacity to observe the emotional reactions of friends *J soc Psychol* 1941 *13* 291–302

403 JOHNSON L H Personality and religious work *Amer J Orthopsychiat* 1932 *2* 317–324

404 JOHNSON T H Personality traits of workers within the field of religion *Relig Educ* 1943 *38* 325–329

405 JOHNSON H M Analysis of Bernreuter's inventory as a predictor of success in certain vocations A problem in scientific method *Psychol Bull* 1941 *38* 8 694 695 (abstract)

406 JOHNSON W The effect of mood on personality traits as measured by Bernreuter *J soc Psychol* 1934 *5* 515–522

407 JOHNSON W B and TIRMAN I M Personality characteristics of happily married unhappily married and divorced persons *Char Pers* 1935 *3* 290–311

408 JOHNSTON J B and WILLIAMSON E G A follow up study of early scho

lastic predictions at the University of Minnesota *Sch and Soc*, 1934, *40*, 1–8

409 JONES, G and GALBRAITH A The interpretation of standardized tests *Sch and Soc*, 1941, *54*, 224–228

410 JONES R D Prediction of teaching efficiency from objective measures *J exper Educ*, 1946 *15*, 85–99

411 JORDAN A M *Educational psychology* New York Henry Holt, 1928

412 JURGENSEN, C E A test for selecting and training industrial typists *Educ Psychol Measmt*, 1942, *2*, 409–426

413 JURGENSEN, C E Extension of the Minnesota Rate of Manipulation Test *J appl Psychol*, 1948, *27*, 161–169

414 KABACK, G R Vocational personalities New York *Teachers College Contributions to Education*, No 924 1946

415 KANDEL I L *Professional aptitude tests in medicine, law and engineering* New York Teachers College of Columbia University, 1940

416 KEANE, F H, and O CONNOR, J A measure of mechanical aptitude *Pers J*, 1927, *6*, 15–24

417 KEFAUVER G N, et al The horizontal organization of secondary education *U S Office of Education Bull* No 17 1932

418 KELLEY T L *Crossroads in the mind of man* Palo Alto Stanford Univ Press, 1928

419 KELLEY, T L *An activity preference test for the classification of service personnel* Washington, D C Dept of Commerce 1946

420 KELBEY E Doctoral project New York Teachers College, Columbia Univ In process

421 KEMP E H and JOHNSON A A Psychological research on bombardier training *III Aviation Psychology Report* No 9 Washington, D C Government Printing Office 1947

422 KENDALL W E The occupational level scale of the Strong Vocational Interest Blank *J appl Psychol* 1947 *31* 283–288

423 KINNEY I B The relationship of certain factors to success in clerical positions *J appl Psychol* 1933 *17*, 55–62

424 KINSEY A C, et al *Sexual behavior in the human male* Philadelphia W B Saunders, 1948

425 KINTER, M *The measurement of artistic abilities* New York Psychological Corporation 1943

426 KIRKPATRICK F H Common sense about tests *Pers J*, 1943 *21*, 277–281

427 KIRKPATRICK F H, and GETTYS I E Validity of the A C E Psychological Examination II *Va Univ Bull* 1945 *45* 120–125

428 KITSON H D *The psychology of vocational adjustment* Philadelphia Lippincott, 1925

429 KITSON, H D Aptitude testing its contribution to vocational guidance *Occupations,* 1934 *12,* 60–64

430 KITSON, H D Creating vocational interest *Occupations* 1942 *20,* 546–571

431 KITSON H D Can we predict vocational success? *Occupations,* 1948 *26,* 539–541

432 KLOPFER B et al Shall the Rorschach method be standardized? *Amer J Orthopsychi* 1949 *9* 514–529

433 KLOPFER B, and KELLEY D M *The Rorschach technique* New York World Book Co 1942

434 KLUGMAN S F Test scores and graduation *Occupations* 1943 *21* 389–393

435 KLUGMAN S F Test scores for clerical aptitude and interest before and after a year of schooling *J genet Psychol* 1944 *65* 89 96

436 KLUGMAN S F Permanence of clerical interest in relation to age and various abilities *J soc Psychol* 1945 *21* 115–120

437 KLUGMAN S F The effect of schooling upon the relationship between clerical aptitude and interests *J genet Psychol* 1945 *66* 255–258

438 KNAUBER A J Testing for art ability *Education* 1934 *55* 219–225

439 KNAUBER A J Construction and standardization of the Knauber Art Tests *Education* 1935 *56* 165–170

440 KOHN H A Achievement and intelligence examinations correlated with each other and with teachers rankings *J genet Psychol* 1938 *52* 433–437

441 KOPP I and TUSSING I The vocational choices of high school students as related to scores on vocational interest inventories *Occupations* 1947 *25* 333–339

442 KORNER A F Origin of impractical or unrealistic vocational goals *J consult Psychol* 1946 *10* 328–334

443 KORNHAUSER A W and SCHULTZ R S Research on the selection of salesmen *J appl Psychol* 1941 *25* 1–5

444 KRAUSS L J The correlation of adjustment and achievement in delinquent boys *Stud Psychol Psychiat Cathol Univ Amer* No 2, 1941

445 KRUEGER, W C F Note concerning group influence upon Otis S A Test scores *J educ Psychol* 1936 *27* 554–555

446 KUDER G F *Manual to the Kuder Preference Record* Chicago Science Research Associates 1939 1946

447 KUDER G F Note on Classification of items in interest inventories *Occupations,* 1944 *22* 484–487

448 KUHLEN R G Age differences in personality during the adult years *Psychol Bull* 1945 *42* 333–358

449 KURTZ A K Recent research in the selection of life insurance salesmen *J appl Psychol* 1941 *25,* 11–17

450 LALEGER G E Vocational interests of high school girls New York *Teachers College Contributions to Education* No 857, 1942

451 LANDIS C and KATZ S I The validity of certain questions which purport to measure neurotic tendencies *J appl Psychol* 1934 *18* 343–356

452 LANDIS C ZUBIN J and KATZ S Empirical evaluation of three personality adjustment inventories *J educ Psychol* 1935 *26* 321–330

453 LARSON R C Studies in Seashore's Measures of Musical Talent *Univ of Iowa Studies* No 6 1940

454 LAWSHE C H SEMANIK I A and LIFTIN J The Purdue Mechanical Adaptability Test *J appl Psychol* 1946 *30* 442–453

455 LAYCOCK S R The Bernreuter Personality Inventory in the selection of teachers *Educ Admin Sup* 1944 *30* 59–64

456 LAYCOCK S R and HUTCHISON N B A preliminary investigation into the problem of measuring engineering aptitude *J educ Psychol* 1939 *30* 280–289

457 LAZARSFELD P F, and GALDET H Who gets a job? *Sociometry* 1941 *4*, 64–77

458 LEARNED W S and WOOD B D *The student and his knowledge* New York Carnegie Foundation for the Advancement of Teaching 1938

459 LEFLER D W et al Relation of test scores to age and education for adult workers *Educ Psychol Measmt* 1946 *6* 351–360

460 LEIFFT L S Students plans interests and achievement Masters thesis Clark University 1940

461 LEHMAN H C and WITTY P A Constancy of vocational interest *Peds J* 1929 *8* 153–165

462 LEHMAN H C and WITTY P A A second study of play activities in relation to school progress *Social Forces* 1930 *8* 409–415

463 LEHMAN H C and WITTY P A Vocational counseling—the interest inventory *Amer J Psychol* 1932 *44* 801–805

464 LEHMAN, H C and WITTY P A Vocational guidance some basic considerations *J educ Sociol* 1934 *8* 174–184

465 LEHMAN R G Interpretation of the Kuder Preference Record for college students of home economics *Educ Psychol Measmt* 1944 *4* 217–223

466 LENNON R T and BAXTER B Predictable aspects of clerical work *J appl Psychol*, 1945 *29* 1–13

467 LEPLEY, W M (Ed) Psychological research in the theaters of war *AAF Aviation Psychology Report* No 17 Washington D C Government Printing Office 1947

468 LEVERING C W Minnesota Multiphasic Personality Inventory *War Med* 1943 *4* 618–629

469 LEWIS J A Kuder Preference Record and MMPI scores for two occupational groups *J consult Psychol* 1947 *11*, 194–201

470 LIGON E M Administration of group tests *Educ Psychol Measmt*, 1942, 2 387–400

471 LIKERT R A technique for the measurement of attitudes *Arch Psychol*, No 140, 1932

472 LIKERT R A multiple choice revision of the Minnesota Paper Form Board test *Psychol Bull* 1944 41 674

473 LINDGREN H C A study of certain aspects of the Lee Thorpe Occupational Interest Inventory *J educ Psychol* 1947 38 353–362

474 LINDQUIST E L (Ed) *Measurement handbook* Washington D C American Council on Education in press

475 LINE W et al The objective measurement of mental stability *J ment Sci* 1945 91 61–106

476 LIVESAY I M Test intelligence and future vocation of high school seniors in Hawaii *J appl Psychol* 1941 25 679–686

477 LIVESAY I M Test intelligence and college expectation of high school seniors in Hawaii *J educ Res* 1942 35 331 37

478 LONG I Relationship between interests and abilities a study of the Strong Vocational Interest Blank and the Lyon Scientific Aptitude Test *J appl Psychol* 1945 29 191–197

479 LONG I and HILL J Additional norms for the Purdue Pegboard *Occupations* 1947 26 160–161

480 LONG W I and LAWSHE C H Jr The effective use of manipulative tests in industry *Psychol Bull* 1947 44 130–148

481 LORD A B A survey of four hundred forty nine special class pupils *J educ Res* 1933 27 108 114

482 LORGE I The influence of the test upon the nature of mental decline as a function of age *J educ Psychol* 1936 27 60–110

483 LORGE I and BEAL R D Broad occupational groupings by estimated abilities *Occupations* 1942 21 289 295

484 LORGE I and THORNDIKE L I The value of responses in completions test as indicators of personal traits *J appl Psychol* 1941 25 191–199 200–201

485 LOUGH O M Teachers college students and the Minnesota Multiphasic Personality Inventory *J appl Psychol* 1946 30 241–247

486 LOUGH O M Women students in liberal arts nursing and teaching training curricula and the Minnesota Multiphasic Personality Inventory *J appl Psychol* 1947, 31 437–445

487 LOUTTIT C M *Clinical psychology* New York Harper & Brothers 1947

488 LOVETT R F and RICHARDSON M W Selecting sales personnel *Pers J*, 1934 12, 248–253

489 LURIE W A A study of Spranger's value types by the method of factor analysis *J soc Psychol* 1937 8 17–37

490 LURIE, W A and WEISS A Analyzing vocational adjustment *Occupations*, 1942 *21*, 138–142

491 MCADORY, M The construction and validation of an art test New York *Teachers College Contributions to Education*, No 388, 1929

492 MCCARTHY, T J Personality traits of seminarians *Stud Psychol Cathol Univ Amer*, No 4 1942

493 MCDANIEL, J W, and REYNOLDS W A A study of the use of mechanical aptitude tests with selection of trainees for mechanical occupations *Educ Psychol Measmt* 1941 *1* 191–197

494 MCGEHEE W Prediction of differential achievement in a technological college *J appl Psychol* 1943 *27* 88–92

495 MCGEHEE W Freshman grades and the American Council Psychological Examinations *Sch and Soc* 1938 *47* 222–224

496 MCKINLEY, J C, and HATHAWAY S R A multiphasic personality schedule II A differential study of hypochondriasis *J Psychol*, 1940, *10*, 255–268

497 MCKINLEY J C and HATHAWAY S R A multiphasic personality schedule IV Psychasthenia *J Appl Psychol*, 1942 *26*, 614–624

498 MCLAUGHLIN SISTER MARY AQUINAS The genesis and constancy of ascendance and submission as personality traits *Univ of Iowa Studies in Educ*, No 5, 1931

499 MCMURRY, R N Efficiency work satisfaction and neurotic tendency *Pers J*, 1932, *11*, 201–210

500 MCMURRY, R N, and JOHNSON, J I Development of instruments for selecting and placing factory employees *Adv Mgt* 1945 *10*, 113–120

501 MCNEMAR Q Critical examination of the University of Iowa studies of environmental influences upon the IQ *Psychol Bull* 1940 37 63–92

502 MCNEMAR, Q Opinion—attitude methodology *Psychol Bull*, 1946, *43*, 289–374

503 MACPHAIL A H Q and L scores on the A C E Psychological Examinations *Sch and Soc* 1942 *56* 248–251

504 MACQUARRIE T W A mechanical ability test *J Pers Res*, 1927, *5*, 320–337

505 MANDELL, M, and ADKINS D C Validity of written tests for the selection of administrative personnel *Educ Psychol Measmt*, 1946 *6*, 293–312

506 MANCOLD B J An analysis of the Kuder Preference Record Master's thesis, MacMurray College, Jacksonville, Ill

507 MANUEL H T, et al The new Stanford Binet at the college level *J educ Psychol*, 1940, *31*, 705–709

508 MANZER, C W The effect of self interest on scores made on the Allport test for measuring ascendance submission *Psychol Clinic*, 1933 *21*, 268–270

509 MARGOLIS, B D The problem of "facade" in the counseling of low scholarship students *J consult Psychol*, 1945 9, 138–141

510 MARKT, A R, and GILILAND A R A critical analysis of the George Washington University Teaching Aptitude Test *Educ Admin Sup*, 1929, *15*, 660–666

511 MARSH C J The diagnostic value of the Bell Adjustment Inventory for college women *J soc Psychol*, 1943 *17*, 103–109

512 MARSHALL M V What intelligence quotient is necessary to success? *J higher Educ*, 1943 *14*, 99–100

513 MASON, C W, and CLEETON, G U Measuring executive ability *Pers J*, 1935, *13*, 277–279

514 MATHLWSON S B *Restriction of output among unorganized workers* New York Viking Press 1931

515 MATTHEWS, L H An item analysis of measures of teaching ability *J educ Res*, 1940 *33* 576–580

516 MEEHL P E The dynamics of structured personality tests *J clin Psychol*, 1945 *1* 296–303

517 MEEHL P L, and HATHAWAY, S R The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory *J appl Psychol*, 1940 *30*, 525–564

518 MEIER, N C Aesthetic judgment as a measure of art talent *Univ of Iowa Studies Series on Aims of Research* No 19 1926

519 MEIER, N C Factors in artistic aptitude *Psychol Monog*, No 5, 1939

520 MEIER, N C Recent research in the psychology of art *Yrb nat Soc Stud Educ*, 1941, *40* 379–400

521 MEIER N C *Art in human affairs* New York McGraw Hill 1942

522 MEIER N C and SEASHORE C F The Meier Seashore Art Judgment Test Iowa City Bur of Res and Service Univ of Iowa 1929

523 MELTENBRUCH P L Preliminary report on the Miami Oxford Curve Block Series *J appl Psychol* 1940 *30* 129–134

524 MELTZER H Approach of the clinical psychologist to management relations *J consult Psychol* 1944 *8* 165–174

525 Memoirs The National Academy of Sciences 1921, *15* Ch 15

526 METFESSEL M Personal factors in motion picture writing I Interests and attitudes *J abnorm soc Psychol* 1935 *30* 333–347

527 MICHEL, J C and BUHLER, C Experiences with personality testing in a neuropsychiatric department of a public general hospital *Dis nerv System*, 1945 *6* 205–211

528 MILES, C C Age and certain personality traits of adults *Psychol Bull*, 1933 *30*, 570

529 MILES C C Psychological study of a young male pseudo hermaphrodite

(In McNemar, Q, and Merrill, M A, *Studies in Personality*) New York, McGraw Hill, 1942

530 Miller, D C Personality factors in the morale of college trained adults *Sociometry* 1940 *3* 367–382

531 Miller, D C Economic factors in the morale of college-trained adults *Amer J Sociol* 1941 *47* 149–156

532 Miller, W S Variation and significance of intelligence quotients obtained from group tests *J educ Psychol* 1924 *15* 360

533 Mitrano, A J Relationship between age and test performance of applicants to a technical industrial high school *J appl Psychol*, 1942, *26*, 482–486

534 Moffie, D J Validity of self estimated interest *J appl Psychol*, 1942, *26*, 606–613

535 Moon, C R Study of premedical and medical scholastic records of students in the University of Illinois College of Medicine *J Assn Amer Med Coll* 1938, *13* 208–212

536 Moore, B V Analysis of tests administered to men in engineering defense training courses *J appl Psychol* 1941 *25* 619–635

537 Moore, H Experience with employment tests Natl Indust Conf Bd *Studies in Pers Policy* No 32 1941

538 Moore, H *Psychology for business and industry* New York McGraw Hill, 1942

539 Morgan, C D, and Murray, H H A method for investigating fantasies *Arch Neurol Psychiat* 1935 *34* 289–306

540 Morgan, W J Some remarks and results of aptitude testing in technical and industrial schools *J soc Psychol* 1944 *20* 19–79

541 Morgan, W J Scores on the Revised Minnesota Paper Form Board Test at different grade levels of a technical industrial high school *J genet Psychol*, 1944 *64* 159–162

542 Morris, C M A critical analysis of certain performance tests *J genet Psychol*, 1939 *54* 85–105

543 Morrow, R S An analysis of the relations among tests of musical, artistic, and mechanical abilities *J Psychol* 1938 *5* 259–263

545 Morton, N W *Occupational abilities* New York Oxford Univ Press, 1935

546 Mosier, C I Group of factors in college curricula *J educ Psychol*, 1935, *26*, 518–522

547 Mosier, C I Factors influencing the validity of a scholastic interest test *J educ Psychol*, 1937, *28* 188–196

548 Mosier, C L A critical examination of the concepts of face validity *Educ Psychol Measmt* 1947 *7* 191–206

549 Moss, F A Aptitude tests in selecting medical students *Pers J*, 1931, *10*, 79–94

550 Moss, F A Report of the Committee on Aptitude Tests for Medical Schools *J Assn Amer Med Coll*, 1942 *17*, 312–315

551 Munroe, R L Inspection technique *Rorschach Res Exch*, 1941, *5*, 166–190

552 Munroe, R L Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method *Appl Psychol Monog*, No 7 1945

553 Munrol, R L Rorschach findings on college students *J consult Psychol*, 1946 *10* 301–316

554 Murphy, G *Personality* New York Harper & Brothers 1947

555 Murphy, G Murphy L B and Newcomb T M *Experimental social psychology* New York Harper & Brothers 1947

556 Murphy, L W The relation between mechanical ability tests and verbal and non verbal intelligence tests *J Psychol* 1936 *2*, 354–366

557 Murray H H *Explorations in personality* New York Oxford Univ Press 1938

558 Murray H A and MacKinnon D W Assessment of OSS personnel *J consult Psychol* 1946 *10* 76–80

559 Mursell, J L What about music tests? *Music Educ J*, 1937, *24*

560 Mursell, J L *Psychological testing* New York Longmans Green, 1946

561 Myers C S Occupational interests a critical notice *Occup Psychol*, 1944 *18*, 148–152

562 NDRC Project SOS 7 An activity preference test for the classification of service personnel Cambridge Harvard Univ 1944

563 NDRC Project N 106 Report No 16 An evaluation of the Personal Inventory for predicting success in parachute school Princeton N J College Entrance Examination Board 1945

564 NDRC Project N 113, Report No 10 Final report in summary of research on the Personal Inventory and other tests Providence R I Brown Univ, 1944

565 Neff, M O, and Mathews, C O Needs of superior students *J higher Educ*, 1935, *6*, 29–34

566 Nelson, E and Nelson N Students attitudes and vocational choices *J abnorm soc Psychol*, 1940 *35* 279–282

567 Newman F B The adolescent in social groups *Applied Psychol Monog*, No 9 1946

568 Newman H H Freeman F N and Holzinger C J *Twins a study in heredity* Chicago Univ of Chicago Press, 1937

569 Northby, A S Prediction of scholastic success in the college of engineering

and architecture *Univ of Minnesota Studies in Articulation*, 1937, 42–47

570 NOTTINGHAM, R D A psychological study of forty unmarried mothers *Genet Psychol Monog*, 1937, *19*, 155–288

571 O CONNOR, J *Born that way* Baltimore Williams and Wilkins, 1928

572 OHMANN, O Q A report of research on the selection of salesmen at the Tremco Manufacturing Co *J appl Psychol*, 1941, *25* 18–29

573 OLDLR H J A note on the twenty minute time limit of the Otis S A tests when used with superior high school students *J appl Psychol*, 1942, *26*, 241–244

574 OLDFR, H J An objective test of vocational interests *J appl Psychol*, 1944, *28*, 99–108

575 OLSON W C The waiver of signature in personal reports *J appl Psychol*, 1936 *20*, 442–450

576 OMWAKF, K T, et al Interrelations of certain physiological measures and aspects of personality *Char Pers*, 1934 *3* 64–71

577 OTIS, A S *Manual for the Otis Self Administering Tests of Mental Ability* Yonkers N Y World Book Co, 1922

578 OTIS, A S The selection of mill workers by mental tests *J appl Psychol*, 1920 *4*, 339–341

579 OTIS J L Prediction of success in power sewing machine operating *J appl Psychol* 1938, *22* 350–466

580 OTIS, J L Procedures for the selection of salesmen for a detergent company *J appl Psychol* 1941 *25* 40–49

581 OTIS, J L, and SMITH, K R Job psychographs in job analysis *Occupations* 1934, *12*, 1–54

582 PAGE H L Detecting psychoneurotic tendencies in Army personnel *Psychol Bull* 1945 *42* 615–658

583 PAGE M L The modification of vicendant behavior in pre school children *Univ of Iowa Studies in Child Welfare* No 3 1946

584 PALLISTER, H, and PIERCE W O D Comparison of Bell Adjustment scores for American and Scottish groups *Sociometry* 1939 *2*, 51–72

585 PALMER I Personal qualities of women teachers of physical education *Res Quart Amer Phys Ed Assn* 1944 *4*, 31–48

586 PATERSON D G Vocational interest inventories in selection *Occupations*, 1946 *25*, 152–158

587 PATERSON, D G (Ed) Research studies in individual diagnosis Univ of Minnesota *Bull Empl Stab Res Inst* No 4, 1934

588 PATERSON, D G, et al *The Minnesota mechanical abilities tests* Minneapolis Univ of Minnesota Press 1930

589 PATERSON, D G, and DARLEY, J G *Men, women and jobs* Minneapolis Univ of Minnesota Press 1936

590 PATERSON, D G, SCHNEIDLER, G, and WILLIAMSON, E G *Student guidance techniques* New York McGraw-Hill, 1938

591 PATERSON, D G, et al The Minnesota Occupational Rating Scales Chicago Science Research Associates, 1941

592 PATTERSON C H On the problem of the criterion in prediction studies *J consult Psychol*, 1946, *10*, 277–280

593 PATTERSON, C H Relationship of Bernreuter scores to parent behavior, child behavior, urban-rural residence, and the background factors in 100 normal adult parents *J soc Psychol*, 1946 *24*, 3–49

594 PATTERSON, C H Is psychotherapy dependent upon diagnosis? *Amer Psychol*, 1948, *3*, 155–159

595 PAYNE, A F Sentence completions New York New York Guidance Clinic, 1928

596 PEDERSEN, R H Validity of Bell Adjustment Inventory applied to college women *J Psychol*, 1940, *9*, 227–236

597 PETERS E F Vocational interests as measured by the Strong and Kuder inventories *Sch and Soc*, 1942, *55*, 453–455

598 PETERS R A study of the intercorrelations of personality traits among a group of novices in religious communities *Stud Psychol Psychiat Cath Univ Amer*, No 7, 1942

599 PETERSON, B H Can you predict who ll be able to go to college? *Sch Exec*, 1941, *61*, 26–27

600 PETERSON, C H The relationship of Bernreuter Personality scores to other parent characteristics, including parent child behavior *Psychol Bull*, 1941, *38*, 8, 704–705 (abstract)

601 PETERSON, D A *Factor analysis of the new United States Navy basic classification test battery* Washington, D C Dept of Commerce, 1946

602 PHILLIPS, W S, and GREENE J B A preliminary study of the relationship of age, hobbies and civil status to neuroticism among teachers *J educ Psychol*, 1939, *30*, 440–444

603 PIÉRON, H L'hétérogénéité normale des aptitudes *Année psychol*, 1945, *41–42*, 1–13

604 PINTNER, R *Intelligence testing* New York Henry Holt, 1931

605 PINTNER, R Neurotic tendency and its relation to some other mental traits *Sch and Soc*, 1932, *36*, 765–767

606 PINTNER R Comparison of interests, abilities, and attitudes *J abnorm soc Psychol*, 1933, *27*, 351–357

607 PIOTROWSKI, Z A Use of the Rorschach in vocational selection *J consult Psychol*, 1943, *7*, 97–102

608 PIOTROWSKI, Z A Differences between cases giving valid and invalid personality inventory responses *Ann N Y Acad Sci*, 1946, *46* 633–638

609. POND, M. Occupations, intelligence, age and schooling *Pers J*, 1933, *11*, 373–382

610. POND, M, and BILLS, M A Intelligence and clerical jobs *Pers J*, 1933, *12*, 41–56

611 POTTS, E M Testing prospective nurses *Occupations*, 1945, *23*, 328–334

612 PRITCHETT, C N The Tennessee Valley Authority Doctor's thesis, Univ of Chicago, 1937

613 PROCTOR, W M Intelligence and length of schooling in relation to occupational levels *Sch and Soc*, 1935, *42*, 783–786

614 PROCTOR, W M A 13 year follow-up of high school pupils *Occupations*, 1937, *15*, 306–310

615. PRUETTE, L, and FRYER D Affective factors in vocational maladjustment *Mental Hygiene*, 1923, *7* 102–118

616 Psychological Research Unit No 1 Clinical techniques project Maxwell Field, Ala PRU No 1, 1944 (mimeo)

617 QUASHA, W H and LIKERT, R Revised Minnesota Paper Form Board Test *J educ Psychol*, 1937, *28*, 197–204

618 RABIN, A I The use of the Wechsler Bellevue Scales with normal and abnormal persons *Psychol Bull*, 1945, *42*, 410–422

619 RAINIER, R A Use of tests in guiding student nurses *Amer J Nurs*, 1942, *42*, 679–682

620 RECKTENWALD, L N Grouping of occupations in the Cleeton Vocational Interest Inventory *Occupations*, 1945 *24* 162–164

621 RECKTENWALD, I N Attitudes toward occupations before and after vocational information *Occupations*, 1946 *24* 220–223

622 RECKTENWALD, L N Homogeneity of items in Cleeton s classification *Occupations*, 1947 *25*, 275–277

623 REID, H B The place of the Bernreuter Personality Stenquist Mechanical Aptitude, and Thurstone Vocational Interest Tests in college entrance tests *Psychol Bull*, 1941, *38*, 711 (abstract)

624 REED, H O The midget wiggly block test for mechanical ability *Industr Arts voc Educ*, 1941, *30*, 153–154

625 REINING H, and STROMSEN, K E Mental abilities and personality traits *Pers Admin*, 1942 14–18

626 REMMERS, H H, and SCHELL, J W Testing the O Connor Wiggly Block Test *Pers J*, 1933 *12*, 155–159

627 RHEINGOLD, H L Interpreting mental defect to parents *J consult Psychol*, 1945, *9*, 142–148

628 RICHARDS, T W Genetic emergence of factor specificity *Psychometrika*, 1941, *6*, 37–42

629 RICHARDSON, H D The intelligence quotient and secondary school guidance *Sch Rev*, 1935, *43*, 49–59

630. RICHARDSON, H M Adult leadership scales based on the Bernreuter Personality Inventory *J appl Psychol*, 1948, *32*, 292–303

631 RICHARDSON, H M, and HANAWALT N G Leadership as related to the Bernreuter personality measures I College leadership in extracurricular activities *J soc Psychol*, 1943, *17* 237–249

632 RIGG, M C Relation of college achievement tests to grades and to intelligence *J educ Psychol*, 1939, *30*, 397–400

633 ROBERTS, J R Manual for the Pennsylvania Bimanual Worksample Test Minneapolis Educational Test Bureau, 1943

634 ROBINSON, J B, and BELLOWS, R M Characteristics of successful dental students *Assn college Registr*, 1941, *16*, 109–122

635 ROE, A The personality of artists *Educ Psychol Measmt*, 1946, *6*, 401–408

636 ROE, A A Rorschach study of a group of scientists and technicians *J consult Psychol*, 1946, *10*, 317–327

637 ROETHLISBERGER, F J and DICKSON W J *Management and the worker* Cambridge Harvard Univ Press 1939

638 ROGERS, C R *Clinical treatment of the problem child* Boston Houghton Mifflin 1939

639 ROGERS, C R *Counseling and psychotherapy* Boston Houghton Mifflin, 1942

640 ROGERS, C R Psychometric tests and client centered counseling *Educ Psychol Measmt*, 1946, *6*, 139–144

641 ROGERS C R and WALLEN, J L *Counseling with returned servicemen* Boston Houghton Mifflin, 1945

642 ROHDE, A R Explorations in personality by the sentence completion method *J appl Psychol*, 1946, *30*, 169–181

643 ROFF, J F The measurement of teaching ability study number two *J exp Educ* 1945 *14*, 52–74

644 RORSCHACH, H *Psychodiagnostik* Bern Hans Huber, 1932

645 ROSANOFF, J *Manual of psychiatry and mental hygiene* New York J Wiley and Sons, 1938

646 ROSE F H Training in speech and changes in personality *Quart J Speech*, 1940, *26*, 193–196

647 ROSE, W A comparison of relative interest in occupational groupings and activity interests as measured by the Kuder Preference Record *Occupations*, 1948, *26*, 302–307

648 ROSENFELD, M A, and NEMZEK, C L Long-range prediction of college marks *Sch and Soc*, 1938, *47*, 127–128

649 ROSS C C Should low-rating college freshmen be told their scores in intelligence tests? *Sch and Soc*, 1938, *47*, 678–680

650 Ross, C C *Measurement in today's schools* New York Prentice-Hall, 1947

651 Ross, L W Results of testing machine-tool trainees *Pers J*, 1943, *21*, 363–367

652 Rostker, L E The measurement of teaching ability Study number one *J exp Educ*, 1945, *14*, 6–51

653 Rothe, H F Output rates among butter wrappers I Work curves and their stability *J appl Psychol*, 1946, *30*, 199–211

654 Rothe, H F Output rates among butter wrappers II Frequency distributions and an hypothesis regarding the restriction of output *J appl Psychol*, 1946 *30*, 320–327

655 Rotter, J B, and Willerman, B The incomplete sentences test *J consult Psychol*, 1947, *11*, 43–46

656 Ruch, F L A technique for detecting attempts to fake performance on a self-inventory type of personality test (In McNemar, Q, and Merrill, M A, *Studies in personality*) New York McGraw Hill, 1942

657 Ruch, F L *Report on the validation of the MacQuarrie Test of Mechanical Ability and the California Short Form Test of Mental Maturity, Language Section, in the selection of winch operators* Washington, D C Dept of Commerce, 1946

658 Rundquist, E A, and Sletto R R *Personality in the depression* Minneapolis Univ of Minnesota Press, 1936

659 Rusmore, J T The R G Pegboard Test of Finger Dexterity *J appl Psychol*, 1942 *26*, 523–529

660 Ryan, T A and Johnson, B K Interest scores in the selection of sales men and servicemen occupational vs ability-group keys *J appl Psychol*, 1942, *26*, 543–562

661 Ryans, D D Tentative statement of the relation of persistence test scores to certain personality traits as measured by the Bernreuter Inventory *J genet Psychol*, 1939, *54*, 229–234

662 Saetveit, J Lewis, D and Seashore C E Revision of the Seashore measures of musical talent *Univ Iowa Stud, Series on Aims of Research*, No 65, 1940

663 St Clair W F The relation of scholastic aptitude to "withdrawal" personality *J educ Psychol*, 1939, *30*, 295–302

664 St Clair W F and Seegers J C Certain aspects of the validity of the Bernreuter Personality Inventory *J educ Psychol*, 1937, *28*, 530–540

665 Sandiford, P, et al Forecasting teaching ability *Univ of Toronto Dept of Educ Bull*, No 81, 1937

666 Sanford, R N, et al *Physique, personality, and scholarship* Washington, D C Soc Res Child Devel 1943

667 Sarbin T R, and Anderson, H C Preliminary study of the relation of

measured interest patterns and occupational dissatisfactions *Educ Psychol Measmt*, 1942, 2, 23–36

668 SARBIN, T R, and BERDIE, R F Relation of measured interests to the Allport-Vernon Study of Values *J appl Psychol*, 1940, 24, 287–296

669 SARTAIN, A Q The use of certain standardized tests in the selection of inspectors in an aircraft factory *J consult Psychol*, 1945, 9, 234–235

670 SARTAIN A Q Predicting success in a school of nursing *J appl Psychol*, 1946, 30, 234–240

671 SARTAIN, A Q Relation between scores on certain standard tests and supervisory success in an aircraft factory *J appl Psychol*, 1946, 30, 328–332

672 SARTAIN, A Q A comparison of the new Revised Stanford Binet the Bellevue Scale, and certain group tests of intelligence *J soc Psychol*, 1946, 23, 237–239

673 SATTER, G A *An evaluation of the Personal Inventory and certain other measures in the prediction of submarine officers' evaluation of enlisted men* Washington, D C U S Dept of Commerce, 1946

674 SCHAEFER B R The validity and utility of the Allport Vernon Study of Values test *J abnorm soc Psychol*, 1936 30, 419–422

675 SCHMIDT B G Changes in personal social, and intellectual behavior of children originally classified as feebleminded *Psychol Monog*, No 5, 1946

676 SCHMIDT, H O Test profiles as a diagnostic aid the Minnesota Multiphasic Inventory *J appl Psychol* 1945, 29 115–131

677 SCHMIDT H O, and BILLINGSLEA, F Y Test profiles as a diagnostic aid the Bernreuter Inventory *J abnorm soc Psychol*, 1945, 40, 70–76

678 SCHNEIDLER, G G Grade and age norms for the Minnesota Vocational Test for Clerical Workers *Educ Psychol Measmt*, 1941, 1, 143–156

679 SCHNEIDLER, G G, and BERDIE, R F Educational ability patterns *J educ Psychol*, 1942, 32, 92–103

680 SCHNEIDLER, G G, and BERDIE, R F Educational hierarchies and scholastic survival *J educ Psychol* 1942 33 199–208

681 SCHNEIDLER, G G, and PATERSON, D G Sex differences in clerical aptitude *J educ Psychol*, 1942, 33, 303–309

682 SCHULTZ, I T, and BARNABAS B Testing for leadership in industry *Trans Kans Acad Sci*, 1945, 48 160–164

683 SCHULTZ, R S Standardized tests and statistical procedures in the selection of life insurance sales personnel *J appl Psychol*, 1936, 20, 553–566

684 SCHULTZ, R S Preliminary study of an industrial revision of the revised Minnesota Paper Form Board Test *J appl Psychol*, 1940, 24, 463–467

685 SCOTT, W D, CLOTHIER, R C, MATHEWSON, S B, and SPRIEGEL, W R *Personnel management* New York McGraw Hill, 1926, 1931, 1941

686 SCUDDER, C L, and RAUBENHEIMER, A S Are standardized mechanical aptitude tests valid? *J juv Res*, 1930, *14*, 120–123

687 SEAGOE, M V Permanence of interest in teaching *J educ Res*, 1945, *38*, 678–684

688 SEAGOE, M V Prognostic tests and teaching success *J educ Res*, 1945, *38*, 685–690

689 SEAGOE, M V Prediction of in-service success in teaching *J educ Res*, 1946, *39*, 658–663

690 SEASHORE C E *The psychology of musical talent* New York Silver, Burdett and Co, 1919

691 SEASHORE C E The discovery and guidance of musical talent *Yrbk of the nat Soc Stud Educ*, 1935 *34*, 447–461

692 SEASHORE, C E Objective measures of musical performance *Univ of Iowa Studies in the Psych of Music*, No 4 1937

693 SEASHORE, C E *Psychology of music* New York McGraw Hill, 1939

694 SEASHORE, C E, et al Seashore measures of musical talents Revised Camden, N J Educ Dept, R C A Mfg Co, 1939

695 SEASHORE, H G Validation of the study of values for two vocational groups at the college level *Educ Psychol Measmt*, 1947, 7, 757–764

696 SEASHORE, H G The improvement of performance on the Minnesota Rate of Manipulation Test *J appl Psychol*, 1947 *31*, 254–259

696a SEASHORE, H G The superiority of college students on the Minnesota Rate of Manipulation Test *J appl Psychol* 1947 *31*, 249–259

697 SEASHORE, H G, and BENNETT, G K A test of stenography some preliminary results *Pers Psychol*, 1948 *1*, 197–209

698 SEASHORE, R H, BUXTON, C E, and MCCOLLUM, I N Multiple factorial analysis of five motor skills *Amer J Psychol*, 1940, *53*, 251–259

699 SEDER, M The reliability and validity of the A C E Psychological Examination, 1938 edition *J educ Res*, 1940, *34*, 91–101

700 SEGEL, D Construction and interpretation of differential ability patterns *J exp Educ*, 1934, *3*, 203–287

701 SEGEL, D Prediction of success in college *U S Office of Education Bull*, No 75, 1934

702 SEGEL, D Differential prediction of scholastic success *Sch and Soc*, 1934, *39*, 91–96

703 SEGEL, D and BRINTLE, S L The relation of occupational interest scores as measured by the Strong Interest Blank to achievement test results and college marks in certain college subject groups *J educ Res*, 1934, *27*, 442–445

704 SEGEL, D, and GERBERICH, J R Differential college achievement predicted by the American Council Psychological Examination *J appl Psychol*, 1933, *17*, 637–645

705 SEGEL, D, and PROFFITT, M M Some factors in the adjustment of college students *U S Office of Education Bull*, No 12, 1937

706 SEIPP, E A study of one hundred clients of the Adjustment Service New York *Amer Assn for Adult Educ*, 1935

707 SELOVER, R B A study of the sophomore testing program at the University of Minnesota *J appl Psychol*, 1942 *26*, 296–307, 456–467, 587–593

708 SEYMOUR, H C The counselor and the high school testing program *Educ Psychol Measmt*, 1946 *6*, 73–80

709 SHAFFER, L F *Psychology of adjustment* Boston Houghton Mifflin 1936

710 SHAFFER L F Fear and courage in aerial combat *J consult Psychol*, 1947, *11*, 137–143

711 SHANNER W M Report on Thurstone Tests for Primary Mental Abilities *Educ Rec Bull*, 1939 *27* 54–60

712 SHANNER, W M, and KUDER G F A comparative study of freshman week tests given at the Univ of Chicago *Educ Psychol Measmt*, 1941, *1*, 85–92

713 SHARKEY V J, and DUNLAP J W Study of reliability and validity of the Academic Preference Blank *J educ Psychol*, 1940, *31*, 103–110

714 SHARTLE C L *Occupational information* New York Prentice Hall, 1946

715 SHARTLE, C L, et al Ten years of occupational research, 1934–1944 *Occupations*, 1944, *22*, 431–441

716 SHUMAN J T The value of aptitude tests for supervisory workers in the aircraft engine and propeller industries *J appl Psychol*, 1945, *29*, 185–190

717 SHUMAN, J T Value of aptitude tests for factory workers in the aircraft engine and propeller industries *J appl Psychol*, 1945 *29*, 156–160

718 SHUTTERLY A Is the aptitude test a panacea? *Occupations*, 1941, *20*, 260–262

719 SKODAK, M, and CRISSEY O L Stated vocational aims and Strong interest scores of high school senior girls *J appl Psychol*, 1942, *26*, 64–74

720 SLATER P Some group tests of spatial judgment or practical ability *Occup Psychol*, 1940, *14*, 39–55

721 SMITH, G B Intelligence and extra curriculum activities selected in high school and college *Sch Rev*, 1936 *44* 681–688

722 SMITH, M University student intelligence and occupation of father *Amer sociol Rev*, 1942 *7*, 764–771

723 SMITH, M An urban rural intelligence gradient *Sociol. soc Res*, 1943, *22*, 307–315

724 SMITH, R V Aptitudes and aptitude testing in dentistry *J dent Educ*, 1943 *8*, 55–70

725 SNYDER T A Reporting intelligence test scores to high school pupils *Sch Rev*, 1937 *45*, 105–111

726 SOLLENBERGER, R T Some relationships between the urinary excretion of male hormone by maturing boys and their expressed interests and attitudes *J Psychol*, 1940, *9*, 179–189

727 SPACHE, G Using tests in a small school system *Educ Psychol Measmt*, 1946, *6*, 99–110

728 SPARLING, E Do college students choose vocations wisely? New York *Teachers College Contributions to Education*, No 561, 1933

729 SPAULDING, F T *High school and life* New York McGraw-Hill, 1938

730 SPEAR, G S The vocational interests of engineering and non-engineering students *J Psychol*, 1948, *25*, 357–363

731 SPEARMAN, C *The abilities of man* New York Macmillan, 1927

732 SPEER, C S The use of the Bernreuter Personality Inventory as an aid in the prediction of behavior *J juv Res*, 1936, *20*, 65–69

733 SPENCER, D *The fulcra of conflict* Yonkers N Y World Book Co, 1938

734 SPRANGER, E *Types of men* New York Stechert, 1928

735 Staff, Division of Occupational Analysis W M C Factor analysis of occupational aptitude tests *Educ Psychol Measmt*, 1945, *5*, 147–155

736 Staff Personnel Research Section, A G O The Army General Classification Test *Psychol Bull*, 1945, *42*, 760–768

737 Staff, Personnel Research Section, A G O The Army General Classification Test, with special reference to the construction and standardization of forms 1a and 1b *J educ Psychol*, 1947, *38*, 385–420

738 Staff, Personnel Research Section, A G O Validation of the General Clerical Abilities Test for selection and placement of War Dept civilian personnel *Amer Psychol*, 1947 *2*, 141–144

739 Staff, Psychological Section, Office of the Surgeon, Ft Worth, Texas Psychological activities in the Training Command, AAF *Psychol Bull*, 1945, *42*, 37–54

740 Staff, Test and Research Section Bureau of Naval Personnel Psychological test construction and research in the Bureau of Naval Personnel the Basic Test Battery *Psychol Bull*, 1945, *42*, 561–571, 638–644

741 STAGNER, R Relation of personality to academic aptitude and achievement *J educ Res*, 1933 *26*, 648–660

742 STAGNER, R Validity and reliability of the Bernreuter Personality Inventory *J abnorm soc Psychol*, 1934 *28*, 413–418

743 STAGNER, R *Psychology of personality* New York McGraw-Hill, 1948 (rev ed)

744 STALNAKER J M Primary mental abilities *Sch and Soc*, 1939, *50*, 868–872

745 STALNAKER, J M Personnel placement in the Armed Forces *J appl Psychol*, 1945, *29*, 338–345

746 STANTON, H M Measuring musical talent *Pers J*, 1928, *7*, 286–292

747 STANTON, H M Prognosis of musical achievement *Eastman School of Music Studies in Psychology* 1929 *1*, 1–89

748 STANTON, H M Measurement of musical talent *Univ of Iowa Studies in Music*, 1935

749 STANTON, M B Mechanical ability of deaf children New York *Teachers College Contributions to Education* No 751 1938

750 STEAD, W H, and SHARTLE C L *Occupational counseling techniques* New York American Book Co 1940

751 STEEL, M, BALINSKY B and LANG H A study on the use of a work sample *J appl Psychol*, 1945 *29* 14–21

752 STEFFLRE B The reading difficulty of interest inventories *Occupations*, 1947, *26*, 95–96

753 STEIN, M I A trial with the criteria of the MacQuarrie Mechanical Ability Test *J appl Psychol* 1927 *11* 391–393

754 STEINMETZ H C Measuring ability to fake occupational interest *J appl Psychol*, 1932 *16* 123–130

755 STENQUIST J L Measurements of mechanical ability New York *Teachers College Contributions to Education* No 130 1929

756 STEPHENS E W A comparison of New England norms with national norms on the revised Minnesota Paper Form Board Test *Occupations*, 1945 *24*, 101–104

757 STEVENS S N and WONDLRLIC E F Relationship of number of questions missed on Otis mental tests and ability to handle office detail *J appl Psychol* 1934 *18* 364–368

758 STEWART NAOMI AGCT scores of Army personnel grouped by occupation *Occupations* 1947 *26* 5–41

759 STODDARD G D Iowa Placement Examinations *Univ of Iowa Studies on Educ* No 2 1925

760 STODDARD G D The IQ its ups and downs *Educ Rec* 1939 *20* 44–57

761 STOGDILL E L and THOMAS M L The Bernreuter Personality Inventory as a measure of student adjustment *J soc Psychol* 1938 *9* 299–315

762 STOKES, T M Selection research in a sales organization *J appl Psychol*, 1941, *25* 41–47

763 STONE C I The personality factor in vocational guidance *J abnorm soc Psychol* 1933 *28* 274–275

764 STOPHER L C The freshman testing program *J higher Educ*, 1941 *12*, 159–162

765 STOTT M B Occupational success *Occup Psychol* 1939 *13* 126–140

766 STRANG R *Personal development and guidance in college and secondary school* New York Harper & Brothers 1934

767 STRANG R *Behavior and background of students in college and secondary school* New York Harper & Brothers 1937

768 STRANG R *Counseling techniques in college and secondary school* New York Harper & Brothers, 1937

769 STRANGE, J R, and SARTAIN, A Q Veterans' scores on the Purdue Pegboard *J appl Psychol*, 1948, *32*, 35–40

770 STRONG, E K, JR Vocational guidance of executives *J appl Psychol*, 1927 *11*, 331–347

771 STRONG, E K, JR *Change of interest with age* Palo Alto Stanford Univ Press, 1931

772 STRONG, E K JR Interests and sales ability *Pers J*, 1934, *13*, 204–216

773 STRONG, E K, JR Permanence of vocational interests *J educ Psychol*, 1934 *25*, 336–344

774 STRONG, E K, JR Predictive value of the vocational interest test *J educ Psychol* 1925, *26*, 331–349

775 STRONG, E K, JR *Vocational interests of men and women* Palo Alto Stanford Univ Press 1943

776 STRONG, E K, JR Personnel psychologists at Stanford University *Psychol Bull*, 1944 *41*, 474–489

777 STRONG E K JR. The interests of forest service men *Educ Psychol Measmt*, 1945, *5*, 157–171

778 STRONG, E K, JR Weighted vs unit scales *J educ Psychol*, 1945, *36*, 193–216

779 STRONG, E K, JR Interests of senior and junior public administrators *J appl Psychol*, 1946 *30*, 55–71

780 STRONG, E K, JR Differences in interests among public administrators *J appl Psychol*, 1947 *31* 18–38

781 STRONG, E K JR Norms for graduate school business students on the Minnesota Vocational Test for Clerical Workers *J appl Psychol*, 1947, *31*, 594–600

782 STRONG, E K JR and HANKES E J A note on the Hankes Test Scoring Machine *J appl Psychol*, 1947, *31*, 212–214

783 STUIT, D B A study of the vocational interests of a group of teachers college freshmen *J appl Psychol* 1938 *27*, 527–533

784 STUIT, D B Prediction of scholastic success in a college of medicine *Educ Psychol Measmt*, 1941, *1*, 77–84

785 STUIT, D B *Personnel research and test development in the Bureau of Naval Personnel* Princeton, N J Princeton Univ Press 1947

786 STUIT, D B The effect of the nature of the criterion upon the validity of aptitude tests *Educ Psychol Measmt*, 1947, 7, 671–676

787 STUIT, D B, and HUDSON, H H The relation of primary mental abilities to scholastic success in professional schools *J exp Educ*, 1942, *10*, 179–182

788 STUIT, D B, and LAPP, C J Some factors in physics achievement at the college level *J exp Educ*, 1941, *9*, 251–253

789 STUIT, D B, and WILSON, J T The effect of an increasingly well defined criterion on the prediction of success at Naval Training School (Tactical Radar) *J appl Psychol*, 1946 *30* 614–623

790 SUPER, D E Occupational level and job satisfaction *J appl Psychol*, 1939, *23*, 547–564

791 SUPER, D E *Avocational interest patterns a study in the psychology of avocations* Palo Alto Stanford Univ Press, 1940

792 SUPER D E The A C E Psychological Examination and special abilities *J Psychol*, 1940, *9*, 221–226

793 SUPER, D E *The dynamics of vocational adjustment* New York Harper & Brothers 1942

794 SUPER, D E The Bernreuter Personality Inventory a review of research *Psychol Bull*, 1942, *39*, 94–125

795 SUPER, D E The place of aptitude testing in the public schools *Educ Psychol Measmt*, 1942 *2*, 267–277

796 SUPER D E The cross sectional and developmental methods of vocational diagnosis *Harvard Educ Rev*, 1942 *12* 283–29,

797 SUPER D E Clinical research in the Aviation Psychology Program of the Army Air Forces *Psychol Bull*, 1944 *41* 551–556

798 SUPER, D E Strong's "Vocational Interests of Men and Women a special review *Psychol Bull*, 1945 *42*, 359–470

799 SUPER D E The cyclical use of directive and non directive counseling techniques *Counseling* (Natl Council of YMCA s) 1947 *5*, 2–5

800 SUPER D E Vocational interest and vocational choice *Educ Psychol Measmt* 1947 *7* 375–384

801 SUPER D E The validity of standard and custom built personality inventories in a pilot selection program *Educ Psychol Measmt* 1947 *7*, 735–744

802 SUPER, D E The Kuder Preference Record in vocational diagnosis *J consult Psychol*, 1947 *11*, 184–193

803 SUPER D E, BRAASCH, W F, and SHAY, J B The effect of distractions on test results *J educ Psychol* 1947, *38*, 373–377

804 SUPER, D E and BROPHY, D A The role of the interview in vocational diagnosis *Occupations*, 1941 *19* 323–327

805 SUPER, D E and HADDAD W C The effect of familiarity with an occupational field on a recognition test of vocational interest *J educ Psychol*, 1943 *34*, 103–109

806 SUPER D E, and ROPER, S An objective technique for testing vocational interests *J appl Psychol*, 1941, *25* 487–498

807 SURGENT, L. L. The use of aptitude tests in the selection of radio tube mounters *Psychol Monog*, 1947, *61*, 1–40

808 SWANSON, D E The role of testing in student personnel services at Hamline University *Educ Psychol Measmt*, 1946, *6*, 25–36

809 SWEM, B R "Accounting aptitude and 'home work *Occupations*, 1945, *23*, 218–219

810 SYMONDS P M *Diagnosing personality and conduct* New York Appleton-Century, 1931

811 TAYLOR, F W *Principles of scientific management* New York Harper & Brothers 1915

812 TAYLOR, H C, and RUSSELL, J T The relationship of validity coefficients to the practical effectiveness of tests in selection discussion and tables *J appl Psychol*, 1939 *28* 565–578

813 TAYLOR, K vF Reliability and permanence of vocational interests of adolescents *J exp Educ*, 1942 *10* 81–87

814 TAYLOR, K vF, and CARTER, H D Retest consistency of vocational interest patterns *J consult Psychol* 1942, *6* 95–101

815 TEEGARDEN, L Manipulative performance of young adult applicants at a public employment office Pts I, II *J appl Psychol* 1942, *26*, 633–652, 754–769

816 TEEGARDEN, L Occupational differences in manipulative performance of applicants at a public employment office *J appl Psychol*, 1943, *27*, 416–437

817 TENDLER, A D A preliminary report on a test for emotional insight *J appl Psychol*, 1930 *14*, 123–136

818 TERMAN, L M Vocational successes of intellectually gifted individuals *Occupations*, 1942, *20*, 498–498

819 TERMAN L M et al *Genetic studies of genius The promise of youth* Palo Alto Stanford Univ Press, 1930

820 TERMAN, I M, and MILES, C C *Sex and personality* New York McGraw-Hill, 1936

821 TERMAN, L M and ODEN M The significance of deviates II Status of California gifted group after 16 years *Yrbk nat Soc Stud Educ*, 1940, *39* (I), 67–74

822 TERMAN, L M, and ODEN M The significance of deviates III Correlates of adult achievement in California gifted group *Yrbk nat Soc Stud Educ*, 1940 *39* (I), 74–89

823 TERMAN L M, and ODEN, M H *The gifted child grows up* Palo Alto Stanford Univ Press, 1948

824 THOMPSON, C E Motor and mechanical abilities in professional schools *J appl Psychol*, 1942, *26*, 24–37

825 THOMPSON, C E Personality and interest factors in dental school success *Educ Psychol Measmt*, 1944 *4* 299–306

826 THOMPSON, C E Selecting executives by psychological tests *Educ Psychol Measmt*, 1947, *7*, 773–778

827 THOMPSON, L A, JR and Associates *Interview aids and trade questions for employment offices* New York Harper & Brothers, 1936

828 THORNDIKE, E L *Prediction of vocational success* New York Common wealth Fund, 1931

829 THORNDIKE, E L Rebounds from the target more about the "Prediction of vocational success" *Occupations*, 1935 *13*, 329–333

830 THORNDIKE, E L *Adult interests* New York Macmillan 1935

831 THORNDIKE, E L *Human nature and the social order* New York Macmillan 1940

832 THORNDIKE, R L Constancy of the I Q *Psychol Bull*, 1940, *27*, 167–186

833 THORNDIKE R L (Ed) Research problems and techniques *AAF Aviation Psychology Report*, No 3 Washington D C Government Printing Office, 1947

833a THORNDIKE, R L Personnel selection New York John Wiley, 1949

834 THORPE, L P California Interest Inventory (In Kaplan O J *Encyclopedia of vocational guidance*) New York Philosophical Library 1948

835 THURSTONE, L L Multiple factor analysis *Psychol Rev*, 1931 *38*, 406–427

836 THURSTONE, L L A multiple factor study of vocational interests *Pers J* 1931, *10*, 198–205

837 THURSTONE, L L A vocational interest schedule *Psychol Bull*, 1935 *32* 719

838 THURSTONE, L L *The vectors of mind* Chicago Univ of Chicago Press, 1935

839 THURSTONE, L L Primary mental abilities *Psychometrika Monog*, No 1 1938

840 THURSTONE, L L, and T G and ADKINS D C The 1938 Psychological Examination *Educ Rec*, 1939, *20* 263–300

841 THURSTONE, L L, and CHAVE E J *The measurement of attitudes* Chicago Univ of Chicago Press, 1929

842 THURSTONE, L L and T G *Factorial studies of intelligence* Chicago Chicago Univ Press, 1941

843 THURSTONE, T G Primary mental abilities of children *Educ Psychol Measmt*, 1941, *1*, 105–116

844 TIEBOUT, C, and MEIER, N C Artistic ability and general intelligence *Psychol Monog*, 1936, *48*, 95–125

845 TIFFIN, J, and ASHER, E J The Purdue Pegboard *J appl Psychol*, 1948 *32*, 234–247

846 TIFFIN, J, and GREENLY, A J Employee selection tests for electrical fixture assemblers and radio assemblers *J appl Psychol*, 1939, *33*, 240–263

847 TINKER, M A Speed, power, and level in the Revised Minnesota Paper Form Board Test *J genet Psychol*, 1944, *64*, 93–97

848 TODD, J E Social norms and the behavior of college students New York *Teachers College Contributions to Education*, No 838, 1941

849 TOLL, C H Scholastic aptitude tests in Amherst College *Sch and Soc*, 1928 *27*, 524–528

850 TOMKINS, S A *The Thematic Apperception Test* New York Grune and Stratton, 1948

851 TOWNSEND, A Achievement and interest ratings for independent school boys *Educ Rec Bull*, 1945, *43*, 49–54

852 TOWNSEND, H Academic aptitude and interest ratings for independent school pupils *Educ Rec Bull*, 1945 *44*, 51–58

853 TRABUE, M R Graphic representation of measured characteristics of successful workers *Occupations*, 1934, *12*, 40–45

854 TRABUE, M R Occupational diagnoses for educational programs *J educ Res*, 1936, *29*, 327–331

855 TRAXLER, A E Reliability, constancy and validity of the Otis I Q *J appl Psychol*, 1934, *18*, 241–251

856 TRAXLER, A E Stability of scores on the Primary Mental Abilities Test *Sch and Soc*, 1941, *53*, 255–256

857 TRAXLER, A L The reliability of the Bell Inventories and their correlation with teachers' judgment *J appl Psychol*, 1941, *25*, 672–678

858 TRAXLER, A E What is the satisfactory IQ for admission to college? *Sch and Soc*, 1940, *51*, 462–464

859 TRAXLER, A E Cumulative test records their nature and uses *Educ Psychol Measmt*, 1941, *1*, 323–340

860 TRAXLER A E *Techniques of guidance* New York Harper & Brothers, 1943

861 TRAXLER, A F Some comments on the prediction of differential achievement in a technological college *J appl Psychol* 1943 *27*, 176–179

862 TRAXLER, A L Note on the reliability of the revised Kuder Preference Record *J appl Psychol*, 1943 *27*, 510–511

863 TRAXLER, A E Correlation between mechanical aptitude scores and 'mechanical comprehension' scores *Occupations*, 1943 *22*, 42–43

864 TRAXLER, A E The correlation between two tests of academic aptitude *Sch and Soc*, 1945, *61*, 383–384

865 TRAXLER, A E Some results of the Kuder Preference Record in independent schools *Educ Res Bull*, 1945, *44*, 37–50

866 TRAXLER, A E Evaluation of aptitude and achievement in a guidance program *Educ Psychol Measmt*, 1946, *6*, 3–16

867 TRAXLER, A E, and HILKERT, R N Effect of type of desk on results of machine scored tests *Sch and Soc*, 1942 *56*, 277–279

868 TRAXLER, A E, and MCCALL, W C Some data on the Kuder Preference Record *Educ Psychol Measmt*, 1941, *1* 253–268

869 TREDICK, V D The Thurstone Primary Mental Abilities Test and a battery of vocational guidance tests as predictors of academic success Unpubl thesis, Penn State College, 1941

870 TRIGGS, F O A study of the relation of Kuder Preference Record scores to various other measures *Educ Psychol Measmt*, 1943, *3*, 341–354

871 TRIGGS, F O A further comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blank for Men *J educ Res*, 1944, *37*, 538–544

872 TRIGGS, F O A further comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blank for Women *J educ Res*, 1944 *38*, 193–200

873 TRIGGS F O A study of the relationship of measured interests to measured mechanical aptitudes, personality and vocabulary *Proc Psychol*, 1947, *2*, 296 (abstract)

874 TRIPLETT, R J Interests of commercial students *J abnorm soc Psychol*, 1934, *29* 409–414

875 TROW, W C Phantasy and vocational choice *Occupations*, 1941, *20*, 89–93

876 TUCKMAN, J The correlations between mechanical aptitudes and "mechanical comprehension scores further observations *Occupations*, 1944 *22*, 244–245

877 TUCKMAN, J A comparison of norms for the Minnesota Rate of Manipulation Test *J appl Psychol* 1944 *28* 121–128

878 TUCKMAN, J A study of the reliability of the Minnesota Rate of Manipulation Test *J appl Psychol*, 1944, *28*, 388–392

879 TUCKMAN, J High school student norms—revised Kuder Preference Record *Occupations*, 1944 *23*, 26–32

880 TUCKMAN, J Norms for the MacQuarrie Test for Mechanical Ability for high school students *Occupations* 1946 *25* 94–96

881 TUCKMAN, J Relationship between subjective estimates of personal adjustment and ratings on the Bell Adjustment Inventory *J appl Psychol*, 1946, *30*, 488–492

882 TURNEY, A H, and COLLINS F I An experiment in improving the personality of high school seniors *J educ Psychol*, 1940, *31*, 550–553

883 TURNEY, A H and FLE, MARY The comparative value for junior high school use of five group mental tests *J educ Psychol*, 1939, *24*, 871–879

884 TURNEY, A H, and FEE, MARY An attempt to use the Bell Adjustment Inventory for high school guidance *Sch Rev*, 1936, *44*, 193–198

885 TUSSING, L An investigation of the possibilities of measuring personality

traits with the Strong Vocational Interest Blank *Educ Psychol Measmt*, 1942 *2*, 59–74

886 TYLER H T Evaluating the Bell Adjustment Inventory *Jun Coll J*, 1936, *6*, 353–357

887 UNGAR F W and BURR E T Minimum mental age levels of accomplishment Albany Univ of State of N Y 1931

888 United States Employment Service *Dictionary of occupational titles* Washington D C Government Printing Office 1939

889 VAN DUSEN A C Permanence of vocational interests *J educ Psychol*, 1940 *31*, 401–424

890 VAN DYNE E V Personality traits and friendship formation in adolescent girls *J soc Psychol*, 1940 *12*, 291–303

891 VAN VORST R B An evaluation of test performances of a group of psychopathic delinquents *Psychol Bull* 1943 *50* 583

892 VAUGHN J and KRUG O The analytic character of the Rorschach Inkblot Test *Amer J Orthopsychi* 1938 *8* 220–229

893 VERNIAUD W M Occupational differences in the Minnesota Multiphasic Personality Inventory *J appl Psychol* 1946 *30* 604–614

894 VERNON M D The drives which determine the choice of a career Pt I *Brit J educ Psychol*, 1937 *7*, 302–316 Pt II *Brit J educ Psychol*, 1938 *8* 1–15

895 VERNON M D The relationship of occupation to personality *Brit J Psychol* 1941 *31*, 294–326

896 VERNON P E Educational abilities of training college students *Brit J educ Psychol* 1939 *11* 253–250

897 VERNON P E Research on personnel selection in the Royal Navy and the British Army *Amer Psychol* 1947 *2* 35–51

898 VERNON P E and ALLPORT G W A test for personal values *J abnorm soc Psychol* 1931 *26* 231–248

899 VITELES M S Job specifications and diagnostic tests of job competency *Psychol Clinic* 1922 *14* 83–105

900 VITELES M S Research in the selection of motormen *J Pers Res* 1925 *4* 100–115 178–199

901 VITELES M S *Industrial psychology* New York W W Norton 1932

902 VITELES M S Getting results from a program of testing for sales ability *Market Serv Amer Mgmt Assn* 1941 *45* 18–31

903 VITELES M S The aircraft pilot 5 years of research a summary of outcomes *Psychol Bull* 1945 *42* 489–526

904 VOTAW J F Regression lines for estimating intelligence quotients and American Council Examination scores *J educ Psychol* 1946 *37*, 179–181

905 WADSWORTH, G W Tests prove their worth to a utility *Pers J*, 1935, *14*, 183–187

906 WALKER, H M *Elementary statistical methods* New York Henry Holt, 1943

907 WALLIS, N A study of tests designed to measure art capacities Master's thesis, Florida State College for Women 1930

908 WARD, L B and KIRK S A Studies in the selection of students for a teachers college *J educ Res*, 1942 *35*, 665–672

909 WARNER W L, and LOW J D *The social system of a modern factory* New Haven Yale University Press 1947

910 WARREN H C *Dictionary of psychology* Boston Houghton Mifflin 1934

911 WARREN N D Personality measurement (In Kaplan O J *Encyclopedia of vocational guidance*) New York Philosophical Library 1948

912 WATSON R I The use of the Wechsler Bellevue Scales a supplement *Psychol Bull*, 1946 *43* 61–68

913 WEBER C D Old and new College Board scores and grades of college freshmen *J Amer Assn college Registr*, 1944 *20* 70–74

914 WECHSLER D *The measurement of adult intelligence* Baltimore Williams and Wilkins, 1944 (rev ed)

915 WEINTRAUB R G and SAILEY R L Graduation prospects of an entering freshman *J educ Res*, 1945 *39* 116–126

916 WEIKER E L, and HARRELL T W Predictive value of certain law aptitude tests *Educ Psychol Measmt*, 1942 *2* 201–207

917 WELLMAN B L Our changing concept of intelligence *J consult Psychol*, 1938 *2*, 97–107

918 WELLMAN B L, SKEELS H M and SKODAK M Review of McNemar's critical examination of Iowa studies *Psychol Bull*, 1940 *37* 93–111

919 WHIPPLE G M *A manual of mental and physical tests* Baltimore War wick and York 1910

920 WHIPPLE G M (Ed) Intelligence its nature and nurture *39th Yrbk nat Soc Study of Educ* 1940

921 WHITE R W Interpretation of imaginative productions (Ch 6 In Hunt J McV *Personality and the behavior disorders*) New York Ronald Press 1944

922 WHITELEY, P L The constancy of personal values *J abnorm soc Psychol*, 1938, *33*, 405–408

923 WHITMAN E C A brief test series for manual dexterity *J educ Psychol*, 1925 *16*, 118–123

924 WIENER D N Individual and group forms of the Minnesota Multiphasic Personality Inventory *J consult Psychol*, 1947 *11* 104–106

925 WICKERT, F (Ed) Psychological research on problems of redistribution

AAF Aviation Psychology Report No 14 Washington, D C Government Printing Office 1947

926 WIGHTWICK M D Vocational interest patterns New York *Teachers College Contributions to Education* No 900 1945

927 WILLIAMSON E G A further analysis of the Young Estabrooks Studiousness Scale *J appl Psychol* 1938 22 105

928 WILLIAMSON, E G *How to counsel students* New York McGraw Hill, 1939

929 WILLIAMSON E G, et al The selection of student nurses *J appl Psychol*, 1938 22 119–131

930 WILLIAMSON E G and BORDIN E S Prediction of success in the College of Science Literature and the Arts *Univ Minn Stud Predict schol Ach* 1942 1, 1–32

931 WILLIAMSON E G and DARLEY J G *Student personnel work* New York McGraw Hill 1946

932 WILSON G M and Staff Adapting the Minnesota Rate of Manipulation Test to factory use *J appl Psychol* 1945 29 346–349

933 WILSON J W and CARPENTER L K The need for restandardization of altered tests *Amer Psychol* 1948 3 172 173

934 WILSON M H The self appraisal program in the Philadelphia Junior High Schools *Educ Psychol Measmt* 1946 6 81–92

935 WITTENBORN J R Mechanical ability its nature and measurement I II Manual dexterity *Educ Psychol Measmt* 1945 5 241–260 395–409

936 WITTENBORN J R TRIGGS F O and FEDER D D A comparison of interest measurement by the Kuder Preference Record and the Strong Vocational Interest Blanks for Men and Women *Educ Psychol Measmt*, 1943 3 239–257

937 WOLF R R JR Differential forecasts of achievement and their use in educational counseling *Psychol Monog* No 1 1941

938 WOOLF M D A study of some relationship between home adjustment and the behavior of junior college students *J soc Psychol*, 1943 17 275–286

939 WONDERLIC E F and HOVLAND C I The personnel test a restandardized abridgement of the Otis S A test for business and industrial use *J appl Psychol* 1939 23 685–702

940 WOOD BEN D and TRAXLER A L Use of the Strong Vocational Interest Blank with public accountants *Certified Public Accountant*, July 1946, 4–5

941 WOODY C Guidance implications from measurements of achievements aptitudes and interests Univ of Michigan *Bur educ Ref Res Bull*, No 156 1944

942 WRENN C G Aiding the fit *J higher Educ* 1935 6, 357–363

943 WRENN, C G Intelligence and the vocational choices of college students *Educ Rec*, 1935 *16*, 217–219

944 WRENN, C G, and CRANDALL, E B Behavior ratings and scholarship among college freshmen *J educ Res* 1941 *34*, 259–264

945 WREN, H A Vocational aspiration levels of adults New York *Teachers College Contributions to Education*, No 855 1942

946 WRIGHT, J H, and LAING, D M The time factor in the administration of the Wonderlic Personnel Test *J appl Psychol* 1943, *27* 316–319

947 YAGER J L The influence of emotional tension in the intelligence test scores of children *Amer Psychol*, 1946, *1*, 464 (abstract)

948 YEAGER, T C An analysis of certain traits of selected high school seniors interested in teaching New York *Teachers College Contributions to Education*, No 660, 1935

949 YOUNG, C W, and LSTABROOKS G H Report on the Young Estabrooks Studiousness Scale for use with the Strong Vocational Interest Blank for Men *J educ Psychol* 1937 *28* 176–187

950 YOUNG, K, DROUGHT N, and BIRGSTRESSER J Social and emotional adjustments of freshmen at the University of Wisconsin *Amer Soc Rev*, 1937 *2*, 166–177

951 YUM K S Primary mental abilities and scholastic achievement in the divisional studies of the Univ of Chicago *J appl Psychol*, 1941 *25*, 712–720

952 YUM K S Student preferences in divisional studies and their preferential activities *J Psychol* 1942 *13* 193–200

953 ZELMAN W R The relationship of the traits of the Bernreuter Personality Inventory to academic success *J Amer Assn College Registr*, 1945 *21*, 81–84

954 ZERGA J L Development and use of apparatus tests in industry *J appl Psychol*, 1944 *28*, 199–202

955 ZYVE, D L A test of scientific aptitude *J educ Psychol*, 1927 *18*, 525–546

INDEX

AAF Aviation Cadet Classification Tests, 17 18, 54, 66, 134 223 228 253, 254 333, 348–350, 379 474 477–480 536 543 658–660
AAF Biographical Data Blank 20
AAF General Information Test 477–480
AAF Mechanical Information Test 478, 479
AAF Satisfaction Test 49 487 528–529
AAF Technical Vocabulary Information Test 477, 479
Abel, T M 99
Ability definition, 61, 117–118 *see also* Aptitude
Abstract trait tests 66–67
Achievement *see* Proficiency
Accountants, 167, 168 171 173 174 175 384, 457 461 496 520–521, 541
Adams Lepley Personal Audit, 336, 528
Adams, W M 343 344
Adjutant General's Department, 91 95–97, 124, 128, 129 235 299
Adkins D C xvii 135, 334 397 452
Administration of tests 72–85 as structuring the counseling relationship 557–562 condition of examinees, 84–85 effects of distractions 73–74 following directions 79–81 motivating examinees, 77–78 557, observing behavior 81–84 order of testing 78–79 preparation of materials 75 proctoring, 75–76 space, 71–75 timing, 81
Advertising men, 434
Aesthetic judgment 308–315 definition, 308 educational and vocational significance 313–314 specific tests of 310–315, tests of, 309
Age influence on test scores, 109 115 164–165 166 170–171, 204, 214 217, 226–228, 259 261 276–277, 295 297 306, 311 390–392, 409 466
Allen, E P, 13
Allport, A S Reaction Study Beckman Revision 336
Allport, G W 376 380 381 382 465–471 481 488 655
Allport Vernon Study of Values 356 376 381 382, 387, 400 407, 423 424, 446 465–471, 478 572
Alteneder, L E 401, 425
Altus, W D 144
American Council on Education 114, 138, 151 347
American Council on Education Psychological Examination, xv 70 73, 88 89 101, 106 107, 114–124, 128 131 136, 144 145 175, 193 251, 266, 289 284 301, 303 334 343, 341 354, 356 360, 368, 374, 429, 452 519
American Educational Research Association xvi
American Institute for Research xx, 26, 349 350 369, 379 542
American Management Association xvi
American Psychological Association xvi, 129 350 418
Anderson, E E, 18, 145
Anderson, H C, 41 139 410
Anderson, V V, 100 104
Andrew D M, 161 165, 171, 172, 173, 174 208 240 291 292 541
Appraising individuals 532–546 *see also* Diagnosis
Aptitude definition 58–61 independence 64–65 lists, 63, 361 369–371, methods of measurement 65–71 quantitative 106 115 116, 117 118 121–122 144–145 450–453 verbal 106, 115 116, 117 118, 121–122 141–145 452–453. *see also* Factor analysis Manual dexterities, Mechanical comprehension Perceptual speed Spatial visualization etc
Army Air Forces xx 17 19 20 22 29 41 51 62–67 76 141, 223 250 269, 287, 332 348 378 434 459 474 487 520 530 541
Army Alpha 64 88, 102 106, 109, 110, 124 128 141 266, 301
Army Beta 289 297 301 303
Army General Classification Test, 64, 94–98 106 107 124–132, 289, 284 349 427
Army psychological research British Army, 84 *see also* Adjutant General's Department Army Air Forces AAF Aviation Cadet Classification Tests, Army Alpha, Army Beta Army General Classification Test Naval Aviation Psychology Navy Bureau of Personnel Royal Air Force
Arsenian S 430 466 472
Artistic ability 315–318 definition 315 educational and vocational significance

317, 318, specific tests of 316–318 tests of, 315, *see also* Aesthetic judgment
Artists, 304 308–318 521 526
Assembly workers 102 104 111, 112 169 200, 210–215 219 220 271, 274 275 294, 305 512
Attitudes 500 513–516
Autobiographies, 11
Average, 644
Awiji Y 98

Babcock H 266–267
Baldwin E F 299
Balinsky B 115 195 293
Ball, F J , 137
Barnabas, B 254 255 431 491
Barnes M W 115, 116
Barnette W L 171 181
Barre M F 190 191
Barrett, D M 176 270, 904, 905 914, 125 436, 457
Batteries of tests See Custom built tests Standard batteries
Baumgarten F , 81
Baxter B , 69 108 300
Beck S J 517 518
Beckham A S 99
Bedell, R 149
Behavior observation in testing 81–84
Bell Adjustment Inventory 41 447, 356, 400, 401 483, 487 502 510–513 514
Bell H M 510 511 512
Bellows R M 269, 340, 426
Bennett, C K 254 368 372 493, 515 509
Bennett Mechanical Comprehension Test 60 118, 191 208 223, 224 225 240, 242 246–260 267, 269 279 902 305 311 342–348 368, 370 374, 398 423 458 491
Benton A L 354 503
Berdie, R F 91, 117, 129 304 306, 377 396, 399, 400, 401 402 403, 404 425 426
Bergstresser J L , 513
Bernreuter Personality Inventory 256 352, 356, 483 487 488–499 500 502 504 510 512, 519, 528
Bernreuter R G , xv, 137 178 334 488–491
Berman I R , 131
Billingslea F Y 192
Bills M A 101 352 410 481 482 494
Binet, A 60 369
Bingham W V , xviii, 18 19 59 60 61 81, 99 108, 169 210 223 228 240 263 264 266
Bixler R H , 551–553, 556 561
Bixler V H 551–553
Blackstone Stenography Test, 157–158
Blackstone Typewriting Test 156–157
Blake R R xx, 49 378 477 528
Blakemore A M , 166, 168
Bluett C H 136
Blum M L 102, 111 167 192–200, 204 207–214 270 271
Bochner, R , 517
Bolanovich D J , 458–459
Bombardiers 39, 129
Bookkeepers 162 168, 169 170 171, 174, 175 461 196
Bordin E S , 403 404 446 556 561
Borow H 342
Braasch W F Jr 71
Brigdon H D 4
Brainard Occupational Preference Inventory 108 416
Bransford I I 122 331
Brewer J M 14 556
Brigham C C 312 313
Brush, F N 230, 232 269 293 301, 342 343
Bryan A I , 118 122 493
Buhler C 504
Burgemeister B B 400 466
Burnham P S 116 141 155, 393, 421
Buros O K xvi
Burr F T 98–99
Burtt H E 378
Business machine operators 110, 168 169, 170, 177 178, 206, 213
Buxton C 184

Calhoon G P , 195
California Adolescent Growth Study 39, 400 402 403
California Test Bureau xvi 260 316 478
California Test of Mental Maturity 106, 141
Callis R 155
Campbell R K 125 126
Candee B , 102 111 167 179, 193 196–200 204 208 213
Cantril, H 368
Capwell D F 504
Cardall Test of Practical Judgment, 336
Carpenter E K 72
Carpenters 213 484 196
Carroll H A 313
Carter H D 46 377, 389, 390 392, 396 399 402–404 409 447 478 488
Case conference in individual appraisal 537–548
Case histories illustrative 589–642 Data 590–606 Thomas Stiles 590–592 Marjorie Miller, 592–595 Ralph Sheridan 595–597 Paul Manuelli 597–599 James G Revere 599–601 Ruth Desmond 601–604 James L Johnson 604–606 Appraisals and counseling 606–627 Thomas Stiles 607–609, Marjorie Miller 609–611, Ralph Sheridan 612–613 Paul Manuelli, 613–615, James G Revere 615–620 Ruth Desmond 620–623 James L Johnson 623–627 Follow ups, Thomas Stiles 629–630 Marjorie Miller 630–631 Ralph Sheridan 634–635 Paul Manuelli 635–638 James G Revere 638 Ruth Desmond, 648–649,

James L Johnson, 639–641, Evaluation, 641
Case records and reports *See* Reports of test results
Cattell, R B 98
Cause and effect 650
CAVD Test, 69 109
Cavett J W 347
Center of Psychological Service George Washington University 315, 351, 355
Channing, A , 99
Chaplin, J P 480
Chapman, J A , 159
Charner, H D , 480
Chemists 384 461
Chesney A M 347
Christensen A M 495
Christensen, T E 445
Christiansen C 504
Cincinnati Employment Center 160, 190
Civil Aeronautics Administration 20 318 350
Clark, W A , 511
Classification *See* Selection
Cleeton, G U , 407
Cleeton Mason Vocational Aptitude Examination, 333 331
Cleeton Vocational Interest Inventory 408 446 465 471–473
Clinical techniques, 19–21 336–340 486, 488, 516–527, 533, 536–540, 511–516
Clothier R C , 98 103
Cohen I M , 31 185 186 218 220
College Entrance Examination Board, 76, 108 155 253 318
Combs Arthur 3, 551–552 561
Confidence levels of 650
Congdon N A 472
Construction of tests cross validation 51–52 factor analyses 53–57 job analysis, 29–31 methods 28–57 selection of criteria 32–13 selection of traits 31–32 standardization 45–48 test construction 13–45 656–657 types of tests, 65–71 validation 48–51
Contingency factors, 655–657
Cook, D W 190 191
Cooperative achievement tests 118 121 135, 151–154 176, 253 340, 356 157 468
Cooperative Test Service, 151, 378
Copeland H A , 175
Correlation, 648–654 661–663
Counseling diagnosis in, 2–14 *see also* Diagnosis processes 2 11 556–562 purpose 1 2 sample summaries of 606–627 tests in 1 6, 7–8 11–14 113–114 123–124, 151, 132 145–146 153–154 155 181–184 201–203 214–217, 220 233–235 244–246 258–260 276–278 295 299 306–307 311–315 317–318 328–329 367–368, 375 413–414 440–444, 462–465 469–471, 496–199 508–510, 512–513 515–516 524, 526 527, 532–533 584–588
Counseling (directive), 5 14, 547–550
Counseling (nondirective), diagnosis in, 4, 14 in vocational guidance 3, 551–556 principles 2–3 550–555
Covner B 551–553 556
Cowdery K M , 408 416
Cox J R 221, 222 223, 343
Coxe Orleans Prognosis Test of Teaching Aptitude, 331 355–356
Crawford A B , 116, 135, 141, 155
Crawford J 285
Crawford Spatial Relations Test, xiii 194, 208, 252 285 291, 302
Crider B 317 511
Crissey O L , 456
Crissy W J E 100 156 468
Criterion of success 15 32–13 339 416–417 480, 656 657 administrative acts, 41–42 internal consistency 42 output, 37–39 proficiency 37 ratings, 39–40 self ratings 40–41
Critical ratio 617–618
Crooks W R 108
Crosby R C 157, 459
Cross O L 116
Custom built tests 17 48–50, 190 198 330–357 177–180 528–529

Dabelstein D H , 117, 460, 462–463
Daniel W J 456
Darley John G xx 5 93 189 207, 363, 384 388 397 399 400–404 414 449, 484, 443 511 527 537–550 552 553
Davis F B 50 379 478 479
Davis P C 478
Dearborn, W F 93
Dentists 153, 206, 269 304 440, 511 426, 458
Dietlein I 157
Detroit Mechanical Aptitude Test 303
Diagnosis data needed in 1 7 need for 2–6 pitfalls in 13–14 sample records 607–627 techniques of xviii, 7–14 536–546
Diamond S 119 150
Differential Aptitude Test Battery 134, 284 331 332 358 364, 368–376 542 566, 569
DiMichiel S G 447, 460, 462–463
Distraction effect on test scores 73–74
Dodge A F 196
Douglass H R 209 210 335 346 347 426
Draftsmen 205 206 213, 232, 290 294 461
Drake, R M 321 325
Drought N I 511 513
DuBois P H 128 129
Duffy E 100 165 166, 168
Dunlap Academic Preference Blank, 476
Dunlap J W 379 407, 413
Dunlop, T S 99

Dvorak B J, 273 358–363, 372, 396
Dyer, H S, 155

Ebert, E, 300
Eckert, R E, 1
Edgerton H A, 137
Educational Test Bureau, 187, 188, 190, 191 202 285 288
Educational Testing Service 151, 343, 348, 355 356
Electricians 243, 496
Ellingson M 10
Ellis, A, 501, 503 504, 505, 511
Ellison M L 137
Elwell Fowlkes Bookkeeping Test 159
Engineering and Physical Sciences Aptitude Test, 134 330, 341–342 343
Engineers 122 138 140 153 232, 269 280, 293, 304 305, 341–343 385, 425–426, 434 461, 469 496
Engineers Council for Professional Development 341
Engineers Northwest, 412
Emerson, M R, 267
Estes S G, 118 122 302 303 385
Estimating achievement from tests, 661–663
Eurich A C 313
Evaluation of use of tests, 21–23 *see also* Case histories, illustrative specific tests sections on validity
Evans, M C 456
Executives, 99 100 201 290 294 333–340 383 460 469 475 493 496 520
Experience effect on test scores 87 115 116, 165–166 201 226–228 236 247 259 393–395, 397, 398, 409 466 485 488, 534

Factor analysis in test construction 53–57, 60–61 106–107 primary mental abilities, 62–65, 107 133–142 360–361 368–372 479
Fairbanks R E 99
Farnsworth P R 324, 325, 489
Faubion R 140 269
Fay, P J 327
Fear R A 254 337
Federl D D 511
Fee M 511
Ferguson L W 108 400 489
Ferson, M L 344
Ferson Stoddard Law Aptitude Examination, 330 343–344 345
Findley, W G 312 313
Fischer, R P 511 512
Fisher, V E 4
Fitts P L, 30
Flanagan J C, xx 22 26 29 30 151 348 356, 357 378 379 479
Flemming C W 351
Flemming E G 351
Flory C D, 336 337 338
Foremen 110, 111, 255 256 257 302, 304, 434, 459 494
Forlano G, 102 111, 512
Forrester G, 540
Forster, M C 396
Fosberg I A, 489
Foundrymen, 104 243
Frandsen, A, 458
Fry, D E, 246
Fryer, D, 11, 94, 103 235 239 377, 378, 408

Garretson, O K, 379, 407
Garrett, H E, 64, 86 284, 643
Gaudet, H, 93, 495
General Aptitude Test Battery (USES), 134, 331 332, 358–368 371, 450, 541 542
General Clerical Test, 164
General Educational Development Tests, 154–155 458
George Washington University Nursing Tests, 315–316
Gerberich J R 120
Ghiselli E E, 176, 179 197–200, 271, 305 431 433
Gilbreth L M, 30
Gilger C A, Jr 377
Glass C F 393 394
Glick H N 84 85
Goodfellow L D, 426, 428
Goodman C H 135 138 140 261, 265 270 271 272, 293 436 458–459
Gordon, H P 493 515
Gotham R E 495
Gouch H G, 504
Greene E B, xvii 310, 378, 474
Greenly R J 112
Gregory Academic Interest Inventory, 476
Griffin C H 342
Griffiths G R, 511
Grocers, 498
Grossmark F xxi
Group Tests administration 72–85 *vs* individual tests 70–71, 112–143 145–146 501
Grune and Stratton 516
Guilford J P 53–57 60–63 67 223, 224, 227 369 459 460 471, 478 506, 515, 527 528
Guilford Martin Personnel Inventory, 336 527, 528
Guilford Zimmerman Aptitude Survey, 369

Hadley H D 545
Hahn, M E 462
Hall, O M 301 514
Halperin F, 517
Hammond H P 150
Hampton P, 491
Hankes, E J, 412 413 414
Hinman B, 238, 299
Hanna, J V 1

Hargan, J , 492
Harmon, L R , 505, 536
Harrell, T W , 140, 194 222–224 227 230–235 266–270, 291, 292, 344
Harris, A J , 207, 208, 209, 210 214, 341
Harris D , 466
Harris, R E , 504
Harrison, R , 526
Hartmann, G W , 488
Hartson, L D 89, 109
Harvard Growth Study 93
Harvard University Press, 524
Harvey, O L , 222
Hathaway, S R , 499 562 503
Hay, E N , 101, 102, 110, 166–169 175–178 181, 182
Heathers, G R , 19, 530
Heinz C , 358
Hemphill J K , 29
Henmon Nelson Test of Mental Ability, 105
Hepner H W , 408
Hertz, M R , 519
Hessemer, M 137
Highsmith J A , 326
Hildreth G , 117
Hill J , 218
Hines M , 206 207
Hobbs R N , 378, 477, 531
Holcomb, G W , 266–269 342, 398 423–426
Home economists, 139 232 293 461 462
Hoppock Job Satisfaction Scale 41 110
Hoppock, R 41, 306, 440
Horn, D 385
Horning S D 266 268
Horst P , 654–657 658
Hovland, C I 108 109
Houghton Mifflin Co , 465
Hull C xvii, 15
Humm, D G , 34
Humm Wadsworth Temperament Test, 352 356, 502 510 528
Humphreys L G , 400
Hunt J McV , 481
Hunt, T 346
Hunter, E C , 115
Hutcheon, N B 342 343
Hutchinson, L , 478

Incomplete sentences tests, 339 488 531
Individual tests *vs* group tests 70–71, 142–143, 145–146 501
Inspectors 33 102 104 112, 169 170, 179 195, 198–199 200 201, 212, 255 256, 257, 271, 304, 305
Intelligence, and educational success 87–92, 109, 120–122, 127–130 136–141 and grade completed, 90–91 and employment, 92–93 and grades 90 and job satisfaction, 103–104 113, 123 and occupational level 94–99, 113, 123, 130–131, and status within an occupation 99–103, 110–113, 123 137–141 144–145, and vocational choice, 92, and vocational success, 93–104 constancy 87, 115, 116 curricular differences, 87–88 curricular satisfaction, 91–92, definitions 86 institutional differences, 88–90 nature and role, 86–104, specific tests of 107–146 types of tests, 105–107
Interests age differences, 390–392, 441, 447 and aptitude, 397–400 422–424, 452–453 and educational success, 424–428 457–459 468–469, and job satisfaction 437–440, 462–463, 469, and occupational preference 128–131 159 and occupational stability 438–439 458, and occupational success, 431–436 459–460 469 and personality traits 400–402 424 455, 457 and socio economic status 388–389 as self concept 402–406, avocational differences in 389 community of 392–399 cultural differences in 417–419 definitions 377–380 398 differentiation in women 436–437 461 factor analysis of, 380–383 477 family resemblances in 99, 397 inventories of, 407 476 occupational differences in 383–388 436–437 460–462, 469 origin of 393–400 sex differences 389–390 specific instruments 407–480 stability of 393–395 421 tests of 407, 477–480 types of 380–388 411
Internal consistency as criterion of success 42–43 642 definition 652
Interpretation of test results See Diagnosis Counseling, Selection Prediction Probability
Interpreting test results to clients 5 538 547–563
Interview in counseling 547–563 in diagnosis 8–9 18–21 159–161 336–338, 538 in selection 336 338 39
Inventories 68–69 379
Iowa Legal Aptitude Test 311–315
Iowa Placement Examinations 149–151

Jiel L M 185
Jacobsen F F , 192–195 207–209, 251–253 302 303
Jacobson C F 426
James W 655
Janitors 290 291 191 511
Janney J F 336–338
Jarvie L L 10
Jenkins J G 32 31 36 318
Jewish Vocational Services, 190, 211
Job analysis 29–31
Job satisfaction and clerical aptitude, 181 and intelligence 103–104 113 123 and interest 137–440, 458 462–463, 469 and manual dexterities 201, 214 and mechanical comprehension 233, 244, 258 276 and personality adjustment,

485, 496 508, 512 and spatial visualization, 295 306
Job setters, 102 111, 256, 257 305
Johnson B K 395, 453
Johnson, D G, xxi
Johnson, D L, 111, 199 254, 255
Johnson E H 496
Johnson W, 489
Jones, E S 209 210
Journalists 140, 153
Jung C, 531
Jurgensen, C E 189–192 200 201

Kaback, G R, 519 520–521
Katz S E, 491
Kelley D M, 517
Kelley, T L 62 476
Kelly E L, 20, 351
Kemble, F, 10
Kendall, W E 388
Kent Shakow Form Boards, 283
Kinsey, A C, 391
Kinter M 310 312, 316
Kirk S A, 495
Kirkpatrick F H 102, 111 512
Kitson, H D, xx 377 382 393 556, 651, 655
Klopfer, B, 517 522
Klugman, S F, 165, 270, 393 429
Knauber, A J, 318
Knauber Art Ability Test, 315 318
Kopp T 459
Kornhauser A W, 352
Kuder G F xvi, 46, 60 120, 135–137, 374 379–382 387 389 397 415–405 513
Kuder Preference Record xiii xvi 60 135 252 302 336 387 397 401 407 409, 423 424 487 445–465, 466, 469 471 473, 474 477 508, 543, 572
Kurtz A K 352
Kwalwasser Dykema Music Test 325

Laborers 201 206, 290 294, 496 541
Land D A 488
Laleger G E 436
Landis C 9 191
Lane D xxi
Lang H, 195, 293
Lipp C J, 304
Larson R C, 324 325
Lislett H R, 266 268 269, 312 308 423 425, 426
Lawshe, C H 279
Lawyers 313–345 384 461
Laycock S R 312 313 495
Lazarsfeld, P F, 93 495
Lee E A 407
Lee Thorpe Occupational Interest Inventory 407, 408 446 475–471
Lefever, D W 298
Leffel E S 241 244 398, 423
Lehman H C 376
Lehman, R T 462
Lepley, W M, 545
Leverenz, C W, 504
Lewerenz Test of Fundamental Abilities in Visual Art, 313, 315–318
Lewis J A 462 507
Librarians 153
Likert, R, 284 297 298, 301, 514
Lindgren H C 173 471
Lindquist F F 154
Link H, 289 285 288
Long I 218 398, 425
Lord A B 99
Lorge I, 531
Lough O M, 506
Lurie W A, 377 380–382

MacKinnon D W, 338
MacPhail, A H 120 122
MacQuarrie T W 264 265
MacQuarrie Test for Mechanical Ability, 113, 125 177 178 198 225, 240 241 250 251, 253 255 260–278 289 298, 302 303 342 313, 360, 398, 423
McAdory M 310
McAdory Art Test 309 310 313 317
McCall W C, 389 447
McCarthy J J, 496 512
McCullough, C M, 209 210
McDaniel, J W, 240, 241, 254, 253, 267, 268
McHale, K 378
McKinley J C 499 502, 503
McKnight and McKnight, 471
McLaughlin Sister Mary Aquinas 485
McMurry R N 111 199 254 255 495
McNamara W J 511 527
McNemar Q 87
Maaske R S 395
Machine operators (factory) 102 200 201 206, 209 213 219 220 232, 249, 254 255 256 257 291 305
Machinists, 242 243
Mahler C A 114
Mallett D R 511
Mandell M 431
Mangold B J 457
Manual dexterities 163, 184–200, 308 319, and job satisfaction 201 214 and success on the job 195–200 206, 210–218 219 and success in training 208–210, definition 184–186 occupational differences in 191 192 200–201 206 213–214 tests of 186–187
Marietta Apparatus Co, 225 285
Marsh C J, 511
Marshall I O 1
Mather M D 393 394 426
Mathews C O 493
Mean 614
Mechanical aptitude 221–281 and educational success 229 232, 240, 241–242, 252–254, 264 265 268–270 and interest 229–230 231, 241 252, and job sat-

isfaction, 233 244, 258, 276 and vocational success, 232, 242–243, 254–258, 270–275, definition, 221–224 factor analysis of, 222–224 occupational differences in, 232–235, 237, 243–244 258, 275–276, specific tests of, 224–281, tests of, 224–225
Mechanics, 129, 131, 140–141, 192, 200 201, 206, 208, 213, 232, 243, 270–271, 280, 290, 294 455, 541
Median, 644
Meier, N C, 308, 309, 319
Meier Seashore Art Judgment Test, 138, 139, 317, 318
Meier Art Judgment Test, 309, 310–315 356
Mellenbruch, P L, 284
Meltzer H 36, 546
Merrill, R A, 346, 347
Metfessel, M, 496
Michel J C 504
Michigan Vocabulary Profile Test, 336, 378, 407, 474–476, 479
Middleton W C 327
Miles, C C, 389 390 402 488
Miller, N E, xx, 29 526
Miller, W S 109
Millwrights, 239, 243
Miniature situation tests 66–67
Minnesota Clerical Test xv 46 60, 102, 118, 138, 139 162 164–183, 188, 292, 334 360, 371 371 123
Minnesota Employment Stabilization Research Institute xvi 46, 93 161, 167 173 174 182, 187 189 190 191 192 202 203, 205, 207 213 211 216, 227 229 232 238, 285 288 291 295 297 358 387 450 495 510 512 587
Minnesota Manual Dexterity Test *See* Minnesota Rate of Manipulation Test
Minnesota Mechanical Abilities Project xvi 40 60 112 187 221 225 228 285 288, 297 300
Minnesota Mechanical Assembly Test 139, 194, 222 229 224, 225–235 236, 240 292 313
Minnesota Medical Aptitude Test 347 348
Minnesota Multiphasic Personality Inventory, 70 401 456–457 469 484 487, 499–510, 527 528
Minnesota Occupational Rating Scale 541
Minnesota Paper Form Board, xv 68 69 118–119 138 139, 140 191 209 222, 228, 229 231, 241, 242, 250, 252 255 256 268, 269 279 283 285, 291, 292 293 294 297–307 313 360 368 374 455
Minnesota Personality Scale, 401, 487, 513–516 527, 528
Minnesota Rate of Manipulation Test, 36 184, 185, 186–203 208 213, 220, 231 251, 287 291 302, 343 360
Minnesota Scale for the Survey of Opinions, 400 401, 514, 515
Minnesota Spatial Relations Test, xv, 68, 72, 138, 139, 140, 187, 194, 209, 222, 229, 231, 283, 285–297, 303
Mitrano, A J, 261, 298
Mode, 644
Moffie, D J, 429
Mollenkopf, W G, 480
Moon, C R, 347
Moore, H, 11 100
Morgan, W J, 266–268 304
Morris, C M, 303
Morris, C N xxi, 72 287
Morrow R S 313
Morton N W, 93 495 496
Moster C L, 150
Moss, F A 347
Moss Hunt Nursing Aptitude Test, 345
Moss Medical Aptitude Test, 330 347–348
Moss Test for Ability to Sell 351
Motivation in testing, 77–78 95–96, 119–130 409–411, 488
Multiple Choice Rorschach, 520
Munroe, R L, 119, 519 521, 523
Murphy, G 481
Murphy I W 266 303
Murray H A, 20 338, 486, 521, 526 529
Murray S, xxi
Murray Thematic Apperception Test, 69, 339 484, 486, 487, 524–527
Mursell J L 326, 327
Musical Talents 319–329 and educational success, 325–326 and musical preferences, 327 definition, 319 320, occupational differences, 327 specific tests of, 321–329 tests of 320
Musicians, 325

National Institute for Industrial Psychology, 13 40, 41 343
National Occupational Conference, xviii
National Society for the Study of Education 87
National Teacher Examinations 335, 356
National Vocational Guidance Association xvi
Naval aviation psychology, 31 35 36
Navigators 56
Navy Basic Classification Test Battery, 134
Navy Bureau of Naval Personnel 250 254
Neel, M O 493
Nelson Denny Reading Test 118 121
Northby A S 150
Nurses 122, 206, 270, 345–347 416, 449, 461, 506

Observation of test behavior 81–84
O Connor J, 203 205, 206, 207, Finger Dexterity Test 187 193 195 197, 203–217, 251, 291 302, 310 311 Tweezer Dexterity Test xiv, 193 203–217, 251,

291, 302, 341, 539 Wiggly Block, 83, 208, 284, 291 302, 303, 544
Occupational ability patterns 363–367
Occupational index, 450
Occupational intelligence levels, 94–99 131
Occupational interest patterns 414–415 450
Occupational success, see specific aptitudes, tests, etc
Office clerks, 101, 102 110, 129, 162–183 192, 205 206, 213, 232, 270 273 274 275, 290, 294, 461, 462, 495 496, 506, 541
Office of Strategic Services, 20, 334 338, 339, 487, 529–530
Ohio State Psychological Examination, 105 137
Ohmann, O O 352
Olson W C, 489
O Rourke, L J, 224 235 238 239, 378, 393 Clerical Test 164, 177 Worksamples 195 196, 208, Mechanical Aptitude Test 68, 209, 223, 225, 231, 235–246 247, 253 259, 267, 269, 278, 302 398 423, 477
Otis A S 368
Otis, J L 209 352, 433
Otis Quick Scoring Test of Mental Ability 106 107
Otis Self Administering Test of Mental Ability 68 69 70 89, 100 102, 105, 107–114 117 128 138 139 141 168 175 177 191, 193 210 222 230 240 242 250, 251, 255, 256 258 266 267, 279 289 300 301, 309 334 347
Output as a criterion of success 37–39 177

Packers 102, 111 169, 170 179 191 195 196–199 200, 201, 203, 206 220 271 275 276, 294, 305
Page, H E, 492
Page, M L 485
Paleontologists, 522–525
Pallister H M, 418 510
Palmer I 495
Paper and pencil tests 68
Paterson D G xx, 72 93 112 164 171 221, 224 228, 229 287 293, 351 363 409 419 434 444 446 455 541
Patterson, C H, 4
Payne, A F, 531
Pennsylvania Bi manual Worksample, 187 193
Perceptual speed 62, 63 136–141 162–183 224 and employment 178 and success in clerical training, 176 and success in clerical work 178 177–181, constancy, 164–166, 179 definition 162 163 occupational differences, 169–170, 173–175 tests of 163–164
Performance tests 68
Perry, J D, 354
Personality, 481–531 and educational success 488, 493–494, 505, 511–512, 519, and job satisfaction, 485, 496, 508, 512, and unemployment, 495, and vocational choice, 484–485 and vocational success, 483–486, 494–495, 506, 512, 515, 519–520 definitions 481–483 measures as screening devices, 485, 491, 493, 503–505, 511 519, 526, 528 occupational differences in, 484–485 495–496, 506–508 512 520–524, 526 specific measures of 488–531 tests and inventories of 486–488 traits, 482, 490, 491, 497–498, 501
Peters, E F, 453
Pharmacists 317, 520–521
Phillips, W S 434
Physical condition and test scores, 84–85
Physicians, 140, 347, 348, 384, 385 426, 469
Pierce, W O D, 418, 510
Pilots, aircraft, 17–19, 20, 22–23, 26, 29–30, 38 54, 56, 129 190 219–220, 348–350 379, 434 459 478 520 528, 545–546 657, 658–660
Pintner General Ability Tests 105
Pintner, R 60, 87 88 102, 103, 468, 495
Piotrowski Z A 447, 452
Placement *See* Selection
Plasterers, 243
Plumbers 243
Policemen 292 290 294, 435, 495 541
Pond M, 98, 101, 266
Potts E M 346
Power 69 108 177–183
Power sewing machine operator 169, 170 209 271, 274 275
Prediction 654–663
Pressmen, 304
Pressey Senior Classification and Verification Tests 105, 172 230 290
Primary mental abilities *See* Thurstone, Factor analysis, Aptitudes
Printers 384
Pritchett C H, 235, 238, 242 243
Probable error 649–650
Probability, 647, 657–663
Probst K 503
Proctor, W M, 91 98
Proficiency as a criterion of success, 37, as a predictor of success 147–148 definition 61 tests 147–161
Profiles of test results 31 415 450, 474–476 536, 540–543 561–572, 576–577
Projective tests, 338, 486, 487, 516–527, 531
Pruette, L 103
Psychograph *See* Profiles of test results
Psychological Corporation, The, xvi xix, 104 142 158 164 246 251, 297, 318, 321 331, 332, 341, 346, 364, 369, 372, 499, 513 542 569

Psychological Institute, 235
Psychologists, 350–351 385, 418–419, 427, 433–474
Psychometric index, 450, 536, 543–546
Psychometric profile *See* Profiles of test results
Purdue Mechanical Adaptability Test, 278–281
Purdue Pegboard, 186, 187, 217–221 225 278–281
Purdue University Division of Applied Psychology 278

Quasha W H , 284 297, 298 301
Questionnaires, 9–10

Rabin, A I 143–145
Radio operators, 129
Range, 645
Rating scales in diagnosis 10, 487 as a criterion of success, 39–41
Raubenheimer A S 241, 267 268
Records, 10, *see also* Case histories Profiles, and Reports
Regents Inquiry 1
Relationship *See* Correlation
Reliability 35
Reports of test results, 564–588 as means of understanding findings, 538–539 samples, 578 579–580, 582–584 588 to clients, 584–588, to professionals 564–584
Reynolds, W A , 240, 241, 252, 267 268
Roberts J R 199
Robinson, J B 269 426
Roe A , 119 484 520–523, 526
Rogers, Carl R 2–5 14, 550–553, 556, 557 561
Rohde, A R , 531
Rolfe J F , 101 122
Rorschach H , 516–518
Rorschach Psychodiagnostic Inkblots 52 69, 85, 119 337 339, 447, 484, 486 487 493 516–524 526
Rosanoff, J 499, 501, 502
Rose W 459
Ross L W 209 241 294 303
Rostker, L F 101 122
Rothney, J W M , 93
Rotter J B 531
Royal Air Force, 65
Ruch, F L , 489
Rundquist E A , 514 515
Ryan T A 393, 433

Sales clerks 100 169 170, 201 292 290 294 495 496, 506, 541
Salesmen 22, 100, 290 294, 351, 384 385, 431–433, 455 461 494 496, 507
Salley, R E 120
Sandiford P 495
Sanford R N , 531
Sarbin, T R , 41 400, 439, 440
Sartain, A Q , 102, 111, 112, 240–242, 251–257, 266–271, 301–304, 453 459, 493, 494
Satter, G A 110
Schaefer, B R , 466, 468
Schmidt H W , 492, 504
Schneidler, G G , 117, 164, 165 168, 171, 228
Schultz I T 254, 255 434 494
Schultz R S , 352 494
Schwebel M 317 409
Science Research Associates, xvi 125, 132, 151 154 217, 219 445
Science Talent Search 354
Scientists, 353–355 461 469 522–523
Scores percentile 646 standard 646–647
Scoring tests, 73, 85 412–414 419–420
Scott W D 11 98 103
Scudder C R 241 267, 268
Seagoe M V , 101 122, 456 426 496 512
Seashore C , 310 319 320–324 Measures of Musical Talents 319 321–329 356
Seashore H G , xxi 188, 190–193 368, 372, 569
Seashore, R H , 59, 61 184
Segel D 109, 120, 151, 424
Selecting tests 551–556 561–562
Selection clinical techniques, 18 21, 336–340, 539, 544–546 contrasted with counseling 15–18, 532–538 programs 25–26 339–357 tests in, 15–27 113–114 123–124 131–132 145–146 153–154 157 158 159 181–183 201–203 214–217, 220, 234 245 260, 277–278, 295–297 306–307 315 328–329 330–333, 367–368 375, 410–411, 463–465 499–471, 496–499 508–510 512–513 515–516 524, 526, 527 528–529 532–538
Self concept 402–406
Semanek I A 279
Seminarians, 196 512
Sender Sadie 520
Shaffer I F xx, xxi, 10 20 481
Shanner, W M 120 135–137
Shartle C L 31 52 62 63 160 169, 170 180 199, 212 271 274 275 278, 301 358 362 363 372 510
Shaw J B xxi 74
Shear B xxi
Sheetmetal workers 129, 243
Sheridan Supply Company 476
Shipley Personal Inventory, 528
Shipping clerks 167
Shuman J T 102 110–113 254–260, 304 305
Simmons K , 300
Situation tests, 20 339 529–531
Skill *See* Manual dexterity Proficiency
Skodak M 436
Slater P 229
Sletto, R R 514 515
Smith, I I 299 511
Smith M 116

Smith P , 15
Social Science Research Council, 654, 656
Social workers, 446–447, 460 461–462 507–508
Sollenberger, R T , 591, 402
Sparling, E J 92, 94
Spatial visualization, 17–18 64, 136–141, 224, and educational success, 293–294, 303–304, and job satisfaction, 295, 306, and vocational success, 294, 304–305, definition, 282–284, 292–293, 295, 303, occupational differences in 294–295 305–306, specific tests, 285–307, tests of, 284–285
Spaulding, F T , 1
Spearman, C 62, 292, 364
Speed, 69, 108 141, 177, 183
Spencer, D , 9, 410
Sports and Hobbies Participation Test, 477–480
Spranger, E 380, 382, 465, 466
Stagner, R 481, 493
Stalnaker, J M , 135
Standard batteries of tests, 358–359
Standard deviation, 645–646
Standard error of estimate, 661–663
Standardization 45–48 95, 415–421
Stanford Binet Scale of Intelligence, 105 110, 117, 143, 144
Stanford (7yve) Scientific Aptitude Test, xv 353–354, 398 423
Stanford University Press, 353, 408, 488, 510
Stanton, H M , 320, 322, 326, 327
Stanton, M B 230, 232, 293, 303
Statistics, glossary, 643–654 prediction 654–663
Stead, W H 52 160, 169, 170 180, 199, 212, 271 274 275 278, 305 363
Steel M 195 293 294
Stefflre, B 409 446, 465
Steinmetz H C 67, 410
Stenographers 157–159, 168 176, 206, 425 458
Stenquist Mechanical Aptitude (Picture) Tests, 222 223 224 225 229, 239, 240, 289 292 342 423
Stenquist Mechanical Assembly Test, 224, 225, 229 239 267, 342
Stephens, E W , 299
Stewart, N 94 95, 127, 130 131
Stoddard G D , 87, 149, 150, 344
Stoelting, C H and Co , 203 204
Stokes, T M , 353
Strang R , 10, 11, 39, 87
Strauss, L M 31, 185 186 218 220
Strong E K Jr xvi xx 22, 45, 46, 48 51, 52 129, 170 334, 335, 377, 379–405, 408–444 447 419, 450 451, 458 459 461, 473, 571
Strong F W , 400
Strong Vocational Interest Blank xv xvi, 22 41, 45 48 50, 60 67 241, 252 255–256, 259, 268 342, 347, 351 352, 353, 356, 377, 379–405, 407, 408–444, 445, 446 449, 453–455, 459, 463, 465, 471, 472, 475, 477, 491, 494, 553, 571 572
Stuit, D B , 140, 304, 348 436
Super, D E , 19 29, 74 175 193, 194, 280, 378, 389 477, 479, 526, 528 530
Surgent L V , 218
Swem, B 176
Symonds, P M , 9, 10, 39, 379, 486

Taylor, F W 30
Taylor K V F , 392, 421
Taylor Russell selection-ratio, 272 278
Teachers, 100, 101 122, 205, 206 213, 232, 290, 294, 355–356, 383, 384, 426 434, 460, 461, 462, 469 494–495, 506, 512
Teegarden L 190–194, 202 294
Tellers, 167, 192, 201, 205, 206
Temperament, 485, 499–510, *see also* Personality
Tendler A D 531
Terman, L M 64 91, 389, 402
Terman McNemar Test of Mental Ability, 105 110
Tests *See* topics such as Intelligence Interests Vocational counseling Construction, etc
Thatcher J W , 477
Thompson, C F 100 207–210, 269, 270, 301 304 458 460 475
Thompson L A , Jr , 160
Thorndike, E L , 62 235, 376 531
Thorndike R L xviii, 34 35 37, 87
Thorpe, L P 407
Thurstone, L L , xvi, 114 116 121 122, 123 192 133, 135, 308, 368 370 371, 380–383 514 Surface Development Test, 17, 284 method of factor analysis, 53 61 222 361 Primary Mental Abilities Tests xvi 56 60–63, 107, 132–142, 145, 176, 231, 284, 292, 293, 331–334, 368 373 374, 397, 452, 539 Examination in Clerical Work, 177 Interest Schedule 343
Thurstone T G , 114 116 132, 133
Ticbout C 309
Tiffin, J 112 279
Tinker M 69 300
Toolmakers, 111 256 257, 264, 305
Toops H , 235 239 378
Townsend A , 424
Trixler A E 10 39 74, 117 135 251 252 301 389 447 511
Tredick, V D 198, 199, 176 231 232 293
Triggs, F O , 387 401 402, 449–453, 456 457
Trow W C 377
Truck drivers, 129 201, 239
t Test 647
Tuckman J , 188, 190–193, 241, 263
Turney, A H , 511, 513

Tussing, L , 459
Tyler, H T 511, 512
Typists, 168, 169, 170, 176, 201, 206, 425, 458, 462

Uhlaner, J , 284
Ullman, R R 434
Unger, E W , 98–99
United States Civil Service Commission, 333, 334, 375
United States Employment Service, xvi, 31, 62, 63 67 141, 159, 160, 168 169 170 176, 178, 179 182, 184, 187 193, 194, 199, 200 204 205, 212 214, 271, 273, 292, 293, 305, 351, 358, 359 361, 362, 365, 367 369, 372, 450, 512
United States Public Health Service 350
University of Iowa, Bureau of Educational Research and Service 49 310
University of Minnesota *See* Minnesota Employment Stabilization Research Institute Minnesota Mechanical Abilities Project, Paterson, D G , and Minnesota tests
Up-grading *See* Selection

Validation cross validation, 51–52 items 48–50 scores 50–51
Validity definition 651–654 selection tests 21–23, *see also* Criterion of success, specific tests
Van Dusen, A C 59 61 393 394
Van Vorst R B , 504
Vermaud W M , 506 507
Vernon P E 84 380–382 465–471
Veterans Administration 350
Viteles, M S 91 102 418 352
Vocational adjustment and personal adjustment, 3 as a criterion of success 40, extent of, 1 *see also* Job satisfaction
Vocational Advisory Service of New York, 566, 571
Vocational counselors 460, 462–463
Vocational guidance, 1, 2
Votaw, D F , 117

Wadsworth G W 21
Walker H M 643
Wallace V E xxi
Wallen, J L , 3 29 49 526 528
Wallis, N 317
Ward, L B 495
Ward, L W , 494
Warren, H C , 59
Warren, N D , 482
Watson, R I 143–145
Wechsler Bellevue Scale of Mental Ability, 70, 85, 106, 107, 118, 134, 142–146, 368
Weintraub, R G , 120
Weitz, J , 49 528
Welders, 129 200
Welker, E L , 344
Wellman, B L , 87
Wesman, A , 368, 372 569
West Publishing Co , 545
Whipple, G M , 163
White, R W , 519 531
Whitely P L , 466
Whitman, E C , 203
Wiener, D N 501, 505
Wightwick, M I , 430, 431, 439 440
Willerman B 531
Williams C T 462
Williams and Wilkins, 142
Williamson Edmund G xx, 4, 14, 228, 546 547–550 552, 553
Wilson G M 189
Wilson, J W , 72
Winsor A L 459
Wittenborn J R , 215 216, 222, 223, 227, 230–233, 292
Witty P A 376
Wonderlic F F , 108, 109
Wonderlic Personnel Test, 100, 106, 111, 336
Wood B D 151
Woodworth Personal Data Sheet 486
Woodworth R S , 486 488
World Book Co , xxi 107 156, 157, 159, 355, 474, 476
Wrappers 102, 111, 169 170, 179, 191 195, 196–199, 200, 201, 294, 305
Wrenn C G 92, 155, 430

Yager, J L , 84
Yale J R , xx 125
YMCA secretary, 383, 384 385
YWCA secretaries, 467 469
Young K 511 513
Yum, K S , 136 389

Zelman W R 493
Zones of approximation 661–663