



DELHI UNIVERSITY
LIBRARY

ARTS LIBRARY
(DELHI UNIVERSITY LIBRARY SYSTEM)

Cl No. TU:55.7

G6

Ac. No. 1070333

This book should be returned on or before the date last stamped below. An overdue charge of 10 paise will be collected for each day the book is kept overtime.

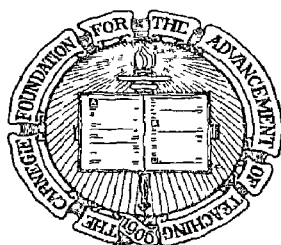
EXAMINATIONS
AND THEIR SUBSTITUTES
IN THE UNITED STATES

By I. L. KANDEL, M.A., Ph.D.

PROFESSOR OF EDUCATION AND
ASSOCIATE IN THE INTERNATIONAL INSTITUTE
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

WITH A PREFACE BY WALTER A. JESSUP
PRESIDENT OF THE FOUNDATION

BULLETIN NUMBER TWENTY-EIGHT



NEW YORK
THE CARNEGIE FOUNDATION
FOR THE ADVANCEMENT OF TEACHING
522 FIFTH AVENUE

1936

D. B. UPDIKE • THE MERRYMOUNT PRESS BOSTON

M. G. Opresco — *Vous avez parlé des examens, Monsieur le Président, et vous avez laissé entendre que vous ne croyiez pas à l'infailibilité des diplômes*

M. Paul Valéry. — *Je ne crois surtout pas à leur valeur durable.*

(LA FORMATION DE L'HOMME MODERNE, Paris)

CONTENTS

	PAGE
PREFACE	vii
INTRODUCTION	xi

CHAPTER I

THE PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING	3
Differences between European and American Education	5
Common Problems in Education	13
The International Examinations Inquiry	19

CHAPTER II

THE TRADITIONAL EXAMINATION	22
Early History	22
Early Criticisms of Examinations	27
Rise of the Accrediting System	34
Admission to College by Certification	35
Establishment of Examination Boards	36
Regents' Examinations in New York State	37
The College Entrance Examination Board	43
Later Criticisms of Examinations and the "Scientific Movement" in Education	57
Experiments in the Reliability of Marking	61
Marking of Essay Examinations	66

CHAPTER III

THE SCIENTIFIC ATTACK ON EXAMINATIONS	69
Selection or Distribution in Education	69
The Scientific Movement in Education	73
The New Type Tests	79
Purposes of Examinations	87
Increasing Enrollments and the Use of Tests	93

Experiments with New Type Tests	95
Columbia College	95
University of Minnesota	100
University of Chicago	103
Statewide Testing Programs	104
Statewide Examinations	107
Ohio	107
Iowa	109
Wisconsin	112
Minnesota	115
Nationwide Testing Programs in the United States	120
College Sophomore Testing Program	121
Coöperative Test Service	126
College Psychological Examination	129
Educational Records Bureau	130
The Pennsylvania Study	133
Cumulative Records and Educational Guidance	140
Cumulative Record Cards	141
Educational Guidance	143

CHAPTER IV

EDUCATION, EXAMINATIONS, AND THE INDIVIDUAL	151
---	-----

APPENDIX

THE INTERNATIONAL EXAMINATIONS INQUIRY	165
INDEX	179

PREFACE

MR. KANDEL's study of *Examinations and their Substitutes in the United States* summarizes the history of a phase of the relationship between our higher and our secondary education which abounds in interest not only for the student of the field but for the lay reader. It is a chronicle of increasing coöperation, of mutual understanding, and, even more pertinent, of growing recognition of the importance of the individual as the product of the American educational process.

A dozen years have passed since Dr. Henry S. Pritchett, then president of the Carnegie Foundation for the Advancement of Teaching, discussed in the Nineteenth Annual Report of the Foundation the direction of our educational progress. Dr. Pritchett pleaded therein for "an ideal of education which rests upon simplicity, sincerity, and thoroughness." This ideal pertains to the quality of education. Now, it has been recognized that in improving the quality of education—an effort in which the Carnegie Foundation has for more than thirty years been concerned—the central problems are three in number: first, the setting up of generally accepted standards of achievement; secondly, the devising of methods of measuring this achievement and of holding pupils to performance; and thirdly, the introduction of such flexibility in educational offerings that each individual may receive the education from which he is able to derive the greatest benefit. The importance of these problems, and especially of their solution, has been repeatedly demonstrated in the study of the relation of secondary and higher education in Pennsylvania, which, in coöperation with agencies in the state, the Carnegie Foundation has been conducting, under the direction of Mr. W. S. Learned, for nearly a decade.

The Pennsylvania study has proceeded simultaneously with another and an even more far-reaching study, namely, the international examinations inquiry, conducted by the International Institute of Teachers College, Columbia University. Funds for both studies have come from Carnegie Corporation of New York in grants to and through the Carnegie Foundation for the Advancement of Teaching. For the international examinations inquiry, under the general direc-

tion of Dr. Paul Monroe, director of the Institute, two international conferences on examinations have been held in England: one at Eastbourne in 1931, and the other at Folkestone in 1935. The results, both provisional and final, of the inquiry in the various countries represented have been published in those lands; Mr. Kandel summarizes them briefly in the appendix to the present volume. From the reports prepared by the national committees one fact stands out with extraordinary clearness, namely: there is an astonishing similarity in some of the educational problems which many countries face. In the United States, because of the unique characteristics of our educational organization and administration, these problems have emerged earlier than in other countries, notably those now represented in the international examinations inquiry. As a result, the questions of individual differences among pupils and the bearing of those differences upon examinations and their use, emerged the earlier in America. These matters began to be subjected to close educational scrutiny and research at least thirty years ago. To this scrutiny the Pennsylvania study will contribute greatly.

In the text of the present Bulletin, Mr. Kandel presents an account of what has been done in American education respecting examinations, the employment of substitutes for their traditional types, and the adaptation of education to individual abilities. He has written this account with the problems which confront other countries in mind. Although he indicates the similarities of these problems he recognizes and takes into account the differences that underlie the educational systems of the various nations of which he writes. Thus the Bulletin presents at once a summary, on the one hand, of what has transpired in the United States respecting examinations and their substitutes, and on the other, of the issues implied for the organization of education in general. These issues appear to be practically the same, the world over. Hence it is hoped that the discussion may have value in the resolution of these and similar issues, not alone in the United States but in other countries.

The thanks of the Foundation are tendered to Carnegie Corporation of New York for making possible not only the international examinations inquiry but the preparation and publication of Mr. Kan-

PREFACE

ix

del's Bulletin; to President Frederick P. Keppel, of the Corporation, for the vision and the lively interest which brought to pass both these inquiries and the volume; and to Dr. Paul Monroe, director of the International Institute of Teachers College, Columbia University, for his talented guidance of the conferences.

WALTER A. JESSUP

October, 1936

INTRODUCTION

THE present volume is one of a series of reports which are being published as the results of an international examinations inquiry, in which educators in a number of European countries and the United States have been coöperating for the past five years.¹ The Inquiry was initiated by Carnegie Corporation of New York in 1931 and has been conducted under the general supervision of Professor Paul Monroe, Director of the International Institute of Teachers College, Columbia University. The first conference on the subject was held at Eastbourne, England, and was attended by representatives from England, France, Germany, Scotland, Switzerland, and the United States. It was clear from the discussions at this Conference, which were published under the title *Conference on Examinations, Eastbourne, England* (New York, 1931), that the problem could only be investigated in the light of educational and social conditions prevailing in each country. A National Committee to undertake such investigations was accordingly organized in each of the countries represented at the Conference. The Reports of these Committees and a Report of the Conference on Examinations, held at Folkestone, England, 1935, at which summaries of these reports were presented and discussed, are now available and are referred to in the text.

Since the problem was not new to the United States and has, indeed, been the subject of research for over a quarter of a century, another investigation was not considered necessary. The present report was, however, undertaken at the request of the Carnegie Corporation to present a unified account of the experience of the United States with examinations and their substitutes in the hope that it will be of service to educators in other countries in which the significance of the problem in its educational and social implications is just beginning to be recognized.

To Dr. Frederick P. Keppel, President of the Carnegie Corporation, the author desires to express his obligation for the opportunity of preparing the

¹ The notion of an international study of examinations had been canvassed for some years before 1931. The place and functions of examinations were discussed at the meeting of the New Education Fellowship at Locarno in 1927, and an international commission was appointed to study the subject in relation to the needs of education. Persons in some twenty countries interested themselves in the matter. At the Elsinore conference, 1929, materials submitted by representatives were discussed. The Fellowship expressed the intention of continuing its inquiry more actively before its conference convened in Nice in 1932. The chief concern of the commission was that the subject of examinations should be studied internationally. The Fellowship published a report in 1935 under the title *Examinations and the Way Out*.

report. To Dr. Walter A. Jessup, President of the Carnegie Foundation for the Advancement of Teaching, the author is indebted for many hours of helpful discussion of the American scene. Finally, the author acknowledges his indebtedness to Professor Ben D. Wood of Columbia University, and to Dr. William S. Learned and Dr. Howard J. Savage for their kindness in reading the manuscript and for many helpful suggestions.

To the following publishers the author is grateful for permission to quote from books published by them: Frederick A. Stokes Company (J. A. Spender, *Through English Eyes*, New York, 1920); Methuen and Company, Ltd. (A. T. Godley, *Lyra Frivola*, London, 1920); Longmans, Green and Company (P. Sandiford, *Educational Psychology*, New York, 1928); The Macmillan Company (D. Starch, *Educational Psychology*, New York, 1920); and Harper and Brothers (R. M. Strang, *Personal Development and Guidance in College and Secondary School*, New York, 1934).

I. L. KANDEL

New York

EXAMINATIONS AND THEIR SUBSTITUTES
IN THE UNITED STATES

CHAPTER I

THE PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

THE world is passing through one of those critical periods of upheaval and unrest which inevitably presage profound and fundamental changes in educational theory and practice. That educational thought and, to a lesser extent, educational practice have already been affected by this unrest is clear to anyone familiar with the literature of education since the outbreak of the Great War.¹ What form education will assume when it passes out of its present stage of transition in such matters as organization, curricula, and methods, is not equally clear. It is quite obvious, however, that, despite the variety of political and economic systems, the educational world is beset by a large number of common problems, not the least important of which is the subject of the International Inquiry on Examinations and of the present volume. And yet the problem of examinations cannot be studied in isolation, nor can it be understood if approached solely from the point of view of techniques of evaluating achievement or of setting educational standards. The proceedings at the first Conference on Examinations,² held at Eastbourne, England, in 1931, revealed that a somewhat complacent acceptance of the infallibility of the examination systems in some of the countries there represented was the expression of a firm conviction in the soundness of the educational organization and aims for which the examinations served as tests. As the national investigations, which were initiated as a result of the conference, progressed in the succeeding four years, that complacency was somewhat shaken. For the investigations, although limited in the main to a study of the reliability of the marking of examinations, have already transcended the purely technical aspects. Properly interpreted the investigations have probed the very roots of the educational systems considered. They have raised the problem of the validity of set patterns of culture or liberal education. They have revealed the existence of wide ranges of individual differences not only among pupils but also among teachers and examiners. They have proved the prevalence of a great amount of educational maladjustment as shown by the percentage of so-called failures in examinations. Where examinations have been used not merely as tests of achievements but for pur-

¹ See I. L. Kandel, *Comparative Education* (Boston, 1933); I. L. Kandel, ed., *Educational Yearbook of the International Institute of Teachers College, Columbia University* (New York, 1925-); Lord Eustace Percy, ed., *Year Book of Education* (London, 1932-).

² See *Conference on Examinations* (New York, 1931).

4 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

poses of selection, they have directed attention to the dangers of injustice through the unreliability of existing systems of marking. And, finally, there is implicit in the results of the investigations a recognition that more is needed than the type of examinations as administered up to the present to secure a complete picture of the abilities and qualities of those who are to be educated. In the crisis which particularly affects youth at present, and perhaps for all time, the evidence available from these studies indicates the necessity of discovering more than can be revealed by the traditional form of examinations, if a sound basis is to be found not merely for educational but for vocational guidance as well. For the changing civilization demands on the one hand an increase and extension of educational opportunities, and on the other the discovery of the right education for the right individual, and beyond this the urgency of devising methods by which there may be reduced the intensity of the current pressure to academic and liberal professions, a pressure which has already been fraught with danger to the stability of society.¹

It is clear then, if one may judge from the results of the examinations up to the present, that a study of the problem of examinations strikes at the very roots of education, whether considered from the cultural or from the social and political points of view. With one exception, the reports of the Scottish Committee, the investigations, although they have in the main not progressed beyond proving the unreliability of marking examinations,² definitely point to the need of studying the question of the distribution and differentiation of education. This does not mean, as may be hastily assumed by some, that the traditional concepts of education and culture have failed; it does mean that, since many pupils who have been exposed to it have not derived the benefits intended to be transmitted, new types of education, better suited to the abilities of the pupils, must be developed. And this is to-day becoming all the more urgent, since there is every indication that the nineteenth-century organization of education with one type of school for the masses and another for the few is about to be surrendered in favor of providing a wider range of facilities for more pupils in post-primary education. In view, then, of such extension upwards the problem of discovering the right education for the right individual is one that not merely concerns teachers and adminis-

¹ The reform of the organization of education at the secondary level in Italy was based on the recognition of this danger; the unrest due to the overcrowding in the universities and consequently in the professions was without doubt a contributory factor to the Nazi Revolution, there is in fact no country in Europe, with the exception of the Soviet Republics, in which this aspect of the educational problem is not serious.

² See Sir Philip Hartog and E. C. Rhodes, *An Examination of Examinations* (London, 1935) and *The Marks of Examiners* (London, 1936); Commission Française pour l'Enquête Carnegie, *La Correction des Épreuves Ecrites dans les Examens* (Paris, 1936).

trators but involves the happiness of individuals and the welfare of society.

In offering in the present volume an account of what has been done on the subject of examinations in the United States, it is not intended to suggest either that a final answer has been found or that the American solution will apply to educational systems in other countries. The discussions at the Eastbourne Conference, which were not formally prepared in advance, brought out no more significant facts than the individuality of national cultures and consequently of their divergent educational problems.¹ Despite these differences, however, the fact that the subject of examinations has been studied in the United States for the past thirty years, that a considerable amount of experimentation has been conducted, and that educational practices have been profoundly affected by the results may have its contribution to make, if only from the point of view of method. On the other hand, educators in the United States may in turn profit from further developments on this problem abroad, for the European countries have on the whole clung more tenaciously to educational traditions, have maintained a higher quality of standards for the few, and have demanded greater thoroughness of achievement.²

DIFFERENCES BETWEEN EUROPEAN AND AMERICAN EDUCATION

Before proceeding to the subject of this volume, then, it is important to discuss certain differences of background in the European countries and the United States in so far as they affect education, particularly since these differences have influenced the approach to the study of the problem of examinations.

The first obvious difference is that of organization. The European organization of education has been and to a large extent still is dual in character—an elementary system for the masses, which even though it may extend beyond the elementary school proper is none the less distinct from the other system of secondary and higher education.³ Hence education above the primary level has had to be protected by barriers in the form of qualifying or

¹ The importance of this point is fully revealed in the *Educational Yearbook, 1929*, of the International Institute of Teachers College, Columbia University, which dealt with "Philosophies underlying National Systems of Education" (New York, 1930); see also I. L. Kandel, *Comparative Education*, Chapters I-IV (Boston, 1933).

² See W. S. Learned, *The Quality of the Educational Process in the United States and Europe*. Carnegie Foundation for the Advancement of Teaching, Bulletin No. 20 (New York, 1927). This excellent comparison of secondary and higher education in the United States and Europe is open to only one criticism: it fails to make any direct reference to the severely competitive character of European education with pressure from the home and pressure for social status, incentives which are on the whole not present in the American scene.

³ See for purposes of illustration the article on France in the *Educational Yearbook, 1930*, pp. 309 ff.

6 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

competitive examinations, a protection which is all the more severe, even though unconscious, partly to maintain the standards of quality and partly to limit the numbers that might ultimately enter the universities and the liberal professions.¹ Examinations accordingly serve a dual purpose; they are used both as tests of achievements and as methods of selection which become more intensive as the time of admission to the universities approaches.

It is obvious, however, that the problem is assuming a new complexion as the old type of organization is beginning to be discarded and the common school idea is being introduced. The traditional academic secondary school is retained in this new scheme, but other types of schools for pupils above the primary stage are gradually beginning to be brought under the general classification of secondary. In the brief period since reforms of this type have been introduced no satisfactory methods of selection have yet been devised. The articulation between the primary and the secondary school proper may have been improved, although this is still doubtful, but the very increase in the number of pupils has only served to intensify the problem of selection and distribution.² It appears to be the general opinion that no satisfactory method of selection has yet been devised, whether by examinations or teachers' judgments, whereby the future success of pupils in the traditional secondary schools can be predicted at the present age of transfer.³ So far as the organization of the educational ladder is concerned one of the crucial points is definitely at the stage where primary education closes and some form of post-primary education begins; the other is at the stage of admission to the universities.

The educational system of the United States is built up on the vertical system, each stage of education being articulated with the next following, so that pupils pass normally on the completion of their elementary education into the high schools and thence into the college under certain conditions, or

¹ The part that the fear of lowering standards and of producing an intellectual proletariat has played in secondary education is well illustrated by the history of secondary education in Germany since the days of Bismarck; it was inherent in the struggle for the recognition of equivalents both in Germany and France; it plays an important part in the conduct of examinations in some of the Latin American countries (see the article on the Argentine Republic in the *Educational Yearbook, 1930*, pp. 1 ff.). The fear of an intellectual proletariat was responsible for the reorganization of secondary education in Italy which definitely represented an attempt to provide differentiated types of secondary schools and to deflect students from entrance to the universities by setting up additional barriers to be overcome (see I. L. Kandel, *Comparative Education*, pp. 757 ff.).

² See C. W. Valentine, *The Reliability of Examinations* (London, 1932), and O. Bobertag, *Schulerausslese* (Berlin, 1935). One can only infer from current articles that the same situation holds in France, where no specific study has yet been made of the effect of the rapidly increasing enrollments in the *lycées* and *collèges*.

³ This takes place at ten years of age in Germany and Italy; at eleven plus in England, France, and the Scandinavian countries.

on the completion of the primary school into the junior high school, thence into the senior high school and from there into the college. There are no entrance examinations barring the way into the high school, and, with few exceptions, students pass from the high school into college under a system of accrediting or certification, provided that certain requirements have been completed in the high school. There is a tendency to reduce the rigidity of these requirements particularly in the state-maintained institutions; and many institutions which have hitherto admitted students on certificates obtained in examinations conducted by the College Entrance Examination Board (see pp. 43 ff. below) have recently agreed to admit students from thirty selected secondary schools, public and private, which have entered upon an eight-year experiment with the secondary school curriculum.¹

The second outstanding difference is to be found in the curriculum of the secondary school. Despite changes which have taken place in the European secondary school curriculum since 1890, chiefly in the recognition of the place of modern languages and sciences, the curriculum is still largely dominated, first, by the traditional concept of a liberal education with its strong emphasis on the study of languages, ancient or modern, and, secondly, by the fact that the secondary school is recognized essentially as preparatory to higher education. This is true even in England, although only about five per cent of the pupils who complete the secondary school course proceed to the universities and where considerable confusion has arisen through the virtual interchangeability of the first secondary school certificate and the matriculation certificate.² What this means in practice in all countries³ is that the schools are organized on the principle that the pupils must be fitted to the curriculum and not the curriculum to the pupils. If a pupil fails in the relevant examinations, then *caveat alumnus*; the standards at least have been maintained!⁴

¹ See the *Reports* of the Commission on the Relation of School and College of the Progressive Education Association.

² In June, 1935, the Northern Universities Joint Matriculation Board announced that these two certificates would be kept distinct and that the changes would go into effect in 1938. This will probably mean the differentiation of courses or requirements for those who plan to take the school certificate examination or the matriculation examination.

³ No study has been made, except in Germany, of the number of pupils who leave school before completing the secondary school course and qualifying for the certificate of admission to the universities. They simply leave with a truncated course.

⁴ Even on this point there no longer exists the complacency which used to prevail. Complaints are made that standards are gradually falling with the increasing numbers who continue to the end of the secondary school and proceed to the universities. A statement to this effect was published in 1928 by the Faculty of Philosophy of the University of Berlin; M. Lapie, who at the time of his death was *Recteur* of the University of Paris, complained of the immaturity of students entering the university and suggested the introduction of a preparatory year. In Denmark a commission was appointed in 1933 to inquire into the same problem,—the lowering of standards consequent on increased numbers.

8 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

The absence (in the United States) of a rigid intellectual tradition explains many of the differences in educational theory and practices noticed by the European observer. While education in Europe is founded on a somewhat traditional and inflexible concept of culture and cultural values, the United States presents a certain cultural rootlessness, despite the criticisms of progressive educators that the high school curriculum is dominated by college entrance requirements. American culture has been the result of two conflicting forces—the frontier and the European tradition, and of these the first has on the whole been the stronger in the greater part of the country.¹ One result of this situation, and, to a certain extent, more intimate and direct control of education by the taxpaying public, has been a greater readiness to experiment with the curriculum at all levels and to adapt its requirements to current social demands and to the needs and capacities of the individual pupil or student. There has thus been a tendency in the past thirty years, slow at first but increasing in momentum in the past decade, to discover courses adapted to the needs, interests, and capacities of the individual, until the only requirements that seem to survive are English and the social studies (history, politics, economics, current events).² The European principle which may be stated as follows: "Find out what the pupil ought to do and make him do it" seems to have been converted in the United States into: "Find out what the pupil can do, and allow him to do it." This means a wider expansion of the curriculum than has ever been thought possible in the European schools, but it also means that methods have had to be devised to discover different types of abilities and aptitudes, and attempts have had to be made to provide such arrangement of courses as will enable a pupil to work to the best of his capacities.³

¹ The predominance of the frontier influence is illustrated in the statement, which is frequently repeated as a justification of the American educational scene, that "we are a pioneer people."

² See I. L. Kandel, *History of Secondary Education*, pp. 481 ff. (Boston, 1930).

³ A sympathetic interpretation of the American situation has been given by J. A. Spender (*Through English Eyes*, pp. 227 f., New York, 1928) "The English academic who has been brought up on Newman's 'Idea of a University,' feels a certain dreariness of spirit, as Matthew Arnold used to say, when he sees the great University of Harvard launching out into an immense and imposing Business faculty. And he laughs sarcastically when he learns that certain American universities of the modern type give degrees and diplomas in horticulture, domestic science, salesmanship, drugstore management, etc. The Englishman, nevertheless, is wrong. This is the natural way in which a new and developing country breaks into education. In such a country the higher education will remain in the air—the privilege of a few favorites of fortune—if it is jealously guarded as the region of a few select studies. A new American university regards it as its first task to go out into the highways and hedges and compel them to come in. In this initial stage its feast must be spread to tempt the appetite; and it follows the universal habit in this business country of discovering what the public wants and giving it without stint or shame. Very unlike Oxford or Cambridge, no doubt, but it is absurd to compare this breaking of ground in a new country with the traditional life of those ancient institutions. Even they are finding that they have

This abandonment of an educational bed of Procrustes, which is still characteristic of European education, has been brought about not by the more flexible concept of culture alone. Three other factors have played an important part in producing this result. First, the traditional justification of "hard" subjects, that they train the mind, was proved early in the century to have validity only under certain conditions; in other words, formal discipline and transfer do not take place automatically, as was assumed for centuries. With the rapid development of high schools, which began about 1905, it was simpler to abandon the whole doctrine rather than to retain it with the reservations which were made.¹ The second factor, which further differentiated American secondary and college education from the European, was the statistical proof of the existence of wide ranges of individual differences; it was found that large numbers of pupils either failed or left school because they were discouraged by failure to meet the requirements of a somewhat narrowly prescribed curriculum. The third factor was the demand of a public, accustomed to more direct participation in educational affairs and less influenced by respect for educational and social traditions than is usual in Europe, that the schools meet the actual and immediate needs and abilities of their children. This demand was rationalized into a theory which was critical of "deferred values" in education and insisted on the right of pupils to an education which is immediate and efficient. Again, the resultant of these factors was to stimulate the study of the individual and the discovery of the right education for him. There is an absence of that bureaucratic control which is generally characteristic of European countries and which inevitably retards the tempo and momentum of change in education, in the United States administrators tend to be more amenable to the demands of the public which supports the school with the consequence that education must be "sold" to it,² which means in turn that it must be adapted to its professed demands. Further, while the European educator is inclined to deplore the lowering of standards due to increased enrollments, the American educator has seized upon these increases, which have been far higher than anywhere else in the

to compromise with new ways, and the new country inevitably starts with them." This statement applies equally, *mutatis mutandis*, to the high schools. On the American background and its effect on education see I. L. Kandel, "Philosophy of Education underlying the System of Education in the United States" in the *Educational Yearbook, 1929*, pp. 461 ff. (New York, 1930); I. L. Kandel, "The American Spirit in Education" in *Essays in Comparative Education*, pp. 49 ff. (New York, 1930); and L. M. Wilson and I. L. Kandel, *Introduction to the Study of American Education*, Ch. I (New York, 1934).

¹ For a reinterpretation of the doctrine see P. Orata, *The Theory of Identical Elements* (Columbus, O., 1928).

² See I. L. Kandel, *Educational Yearbook, 1929*, pp. 500 ff. (New York, 1930).

10 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

world, as the opportunity for varying and differentiating curricula to meet the wider range of individual differences.¹ To put this in another way, while the European secondary schools and universities are intended for those who are presumed to be capable of profiting from their courses, the American high schools and colleges attempt to provide educational programs from which the tested abilities and aptitudes of the students show promise of profiting.

Up to this point the terms "high school" and "secondary school" have been used interchangeably. It must be remembered, however, that the American high school is the school for all adolescents, the only post-primary school, in which the curriculum traditionally associated with the European secondary school constitutes but one of many differentiated curricula offered. With few exceptions the American high school is of the cosmopolitan or multiple bias type, offering in one building the whole range of courses distributed in Europe in secondary schools proper, trade schools, industrial schools, technical schools, commercial schools, and household arts or domestic science schools. As a result, although admission to the high school is not on a selective basis, the problem of adaptation to individual differences of aptitudes has placed the problem of distribution of pupils within rather than at the threshold of the school as in the European school systems. One method which has been employed has been the introduction of the junior high school, which has lowered the age of beginning post-primary education to twelve instead of the traditional fourteen, and has provided for a series of exploratory courses through which pupils could discover their major curricular interests.² Despite this fact, however, it has been found that pupils or their parents tend in the main to elect the academic courses, either because these are still regarded as more respectable or because they lead more directly into college. Whatever the reason, it is now generally recognized that, despite the exploratory courses provided in the junior high schools, too many pupils choose courses for which they are not fitted and that the problem of suitable methods of distribution is more pressing than ever. The situation has, in other words, brought to the fore the problem of discovering more suitable and accurate methods of guidance than are provided by the traditional practice of examinations.³ If the function of education is to discover the abilities and apti-

¹ The crisis in secondary and higher education which is prevalent to-day in all countries is undoubtedly due to an increase in the number of students, more highly differentiated in abilities than ever before in the history of these educational levels.

² See T. H. Briggs, *Junior High School* (Boston, 1920); L. V. Koos, *The Junior High School* (New York, 1925); W. M. Proctor and U. Ricciardi, *The Junior High School, Its Organization and Administration* (Stanford University, Cal., 1930).

³ See E. L. Thorndike, "The Distribution of Education," *School Review*, May, 1932, pp. 335 ff.; H. D.

cludes of pupils and to arrange for them those courses from which they are most capable of profiting the central problem accordingly comes to be, as it has in many school systems, that of guidance and the development of methods, records, tests which will make such guidance valuable. Nor is this problem confined to the high school, for with the rapid increase in the number of students in the college it has also had to be faced there. It cannot be denied, however, that the problem of guidance is complicated even in the United States by the survival of traditional educational values and the association between education and social status.

The rigid adherence to the tradition of culture, which is characteristic of European education and which is perpetuated by strong social approval on the one hand and official sanctions or privileges on the other, has as its counterpart the protection of a body of highly specialized teachers. The result of this specialization is that on the whole teachers are more interested in their own subjects than in their pupils as individuals,¹ with two further consequences, that they tend to be exacting in the maintenance of standards of achievement in the subjects which they teach, and that they fail too often to see the bearing of these subjects on the general all-round education of the pupils under their charge. To some extent these consequences also flow from the ever-present pressure of external examinations and of requirements² which open the way for the pupils to advancement beyond the secondary school, whether to the universities or to selected occupations. The reforms of secondary education in Germany in the brief Republican era following the War were designed to reduce the pressure of specialist teachers and of the system of examinations; the experiment was, however, all too brief and the teachers themselves too steeped in the tradition to justify any conclusion other than those alleged also from the traditional point of view that standards were being lowered. A somewhat similar unrest developed in English secondary education after the War and concentrated upon criticisms of the examination system, one result of which is that noted in the action of the Joint Matriculation Board of the Northern Universities (see note p. 7). Both in England and Germany (during the Republic) the tendencies of edu-

Richardson, "The Selective Function of the Secondary School," *School Review*, Nov., 1933, pp. 685 ff.; I. L. Kandel, *The Dilemma of Democracy* (Cambridge, Mass., 1933); J. L. Tildsley, *The Mounting Waste of the American Secondary School* (Cambridge, Mass., 1936)

¹ This statement may appear too sweeping if applied to the English schools, and yet, despite the English emphasis on the formation of character as the ultimate aim of education, it is still true that English teachers tend to look upon their pupils differently in the classroom and outside of it.

² *Sanctions* in France and *Berechtigungen* in Germany.

12 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

cational thought have moved in the direction of better adaptation of curricula to individual aptitudes.¹

By contrast the teachers in American high schools are not so highly specialized as in the European secondary schools. With a few notable exceptions the requirements for certificates to teach in high schools demand graduation from a college without special reference to subject matter specialization but requiring a certain amount of preparation in professional subjects. This is particularly true in the large number of small high schools where teachers may be called upon to teach a variety of unintegrated subjects. There is some truth in the criticism which is frequently made that more reliance is placed in the preparation of teachers on methods and techniques than on mastery of content.²

The difference between European and American practice may, however, be rationalized in the statement that, while the European teacher is a master of the subjects which he teaches, the American teacher understands his pupils better. One result is that American teachers are more ready to accept innovations; another result which is not wholly due to this cause alone but to a number of the other characteristics of American education and culture is that there is no deep appreciation of or persistence in subjects of study for their own sake.³ The contrast between the European and American teachers also exists in the United States itself, but is not quite so marked, in the contrast between the high school teachers and the teachers in colleges who are more highly specialized in their subjects. But even at the college level, and largely because of the increasing enrollments, there has been a tendency not only to relax the requirements for admission but also to permit greater freedom of choice in the courses taken by students. There is, in other words, the same tendency to consider the aptitudes, capacities, and needs of the individual student rather than to fit him into a preordained educational pattern or mould.

¹ The adoption of "alternative courses" in the secondary schools of the West Riding of Yorkshire and elsewhere may be cited as an example. The Consultative Committee of the English Board of Education is at present conducting an inquiry into the secondary school curriculum.

² This criticism that American practice tends to emphasize training in methods at the expense of mastery of content was the burden of the Report of the League of Nations' Mission of Educational Experts on *The Reorganization of Education in China* (Paris, 1932). The Experts tended to overemphasize this defect because of their lack of familiarity with changes which are proceeding and with the criticisms of the situation current in the United States itself.

³ It is unnecessary at this point to refer to the American "credits," "units" or "points" system accompanied by the system of electives as contrasted with the continuity of a course, once selected, in the European schools. For a history of the fall and imminent decline of the unit system in American education the reader is referred to R. L. Duffus, *Democracy Enters College* (New York, 1936), which also deals with current experiments in a number of colleges to adapt the curriculum requirements to the needs and abilities of the students as individuals.

COMMON PROBLEMS IN EDUCATION

It would, however, be a serious mistake to exaggerate the differences between the educational problems of the United States and the countries of Europe and to minimize the existence of a large number of common problems. A study of the differences does, indeed, bring into relief the problems which to-day equally confront most countries in education. If a science of education, using the term in its broadest sense, is to be developed, much is to be gained from the study of these common problems as they emerge in the varied settings of different national cultures. It is clear that the world is to-day confronted with a new orientation in the education of youth. The nineteenth century was marked by efforts to establish universal compulsory elementary education. The twentieth century has already begun to recognize the inadequacy of elementary education as a preparation for life in the complex world in which we live. These complexities affect every aspect of national and international organization—the economic as much as the political, the social no less than the individual.

On the political side it is realized more than ever before that the mere extension of the suffrage based on a training in literacy alone is no longer adequate; that citizenship in the modern world demands not only a longer and more intensive but a more specific preparation than has been provided hitherto.¹ From the economic point of view the rapid technological developments inevitably mean the unemployability of youth, at least in occupations that conduce to the happiness of the individual and the welfare of society, and the competition that is developing must affect, as it has already affected, both the liberal and other professions. The reduction of the birthrate, except in those countries which deliberately foster its increase for military purposes, means increased provision for the better care and development of youth. From whatever angle the problem is approached, quite clearly a new era is beginning in which society will prolong its educational tutelage of youth. Just as the establishment of universal elementary education was the central problem in the nineteenth, so in the twentieth century there is scarcely a country in the world which is not concerned with the problems both of the education of the adolescent and of higher and professional education.

As long as the social distinctions between elementary education for the masses and secondary education for the select few prevailed the problems involved were not serious. It was taken for granted that the function of the

¹ This need explains the rapid spread of the "social studies" in elementary and high schools and of the greater interest in the social sciences in the colleges and universities of the United States. It is paralleled in England by the recent establishment of the Association for Education in Citizenship.

14 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

elementary school was to impart literacy and the so-called fundamentals of an education and that the purpose of secondary education was to transmit the accepted tradition of a liberal education considered at once as the essential education for leaders and as a preparation for higher education. But even in the field of secondary education the cultural changes of the nineteenth century with the increased recognition of the value and place of sciences and modern languages and the gradual increase in the number of pupils produced an unrest which led to profound modifications in both secondary and higher education.

With the rapid increase in the number of pupils in the secondary schools since the War a new unrest has developed. But the situation, as compared with that in the last two decades of the nineteenth century, has been completely changed. The question to-day is no longer one of revising the curricula and courses of the secondary schools proper, although that is still insistent, but of bringing within the purview of a single system the education of all adolescents. The reorganization in Italy to which reference has already been made, the reform of the educational system in Sweden, and the movements for the *école unique* in France, for the *Einheitsschule* in Germany,¹ and for secondary education for all in England; all these movements are directed to the abolition of the traditional dual organization of education, the adoption of a common foundation of primary education up to the threshold of adolescence, and the creation of alternative types of schools, all of which together with the traditional secondary schools, would constitute education at the secondary level.

It is not necessary here to discuss at length the reasons for these movements, but some of them at least have a bearing on the problems to be considered in this volume. The extension of education is, of course, the result of the recognition of the inadequacy of elementary education as a preparation for life in society to-day; it is equally the result of the inability of adolescents to find suitable employment. Beyond this, however, the extension of educational opportunities is due to a realization that ability and intelligence are not the monopoly of any special stratum in society, and that in the interests of national welfare alone it is desirable to utilize all the intellectual resources that are available entirely irrespective of class. From the social point of view it is also recognized that the results of educational maladjustments, the attempt to fit square pegs into round holes, may lead to social maladjustment

¹ The German movement was abortive except for the establishment of a common *Grundschule* for the first four years of elementary education. The education of the adolescent is at present the subject of inquiry in South Africa and Australia, and in many of the Latin American countries. See in general "The Expansion of Secondary Education" in the *Educational Yearbook, 1930* (New York, 1931).

and in turn prove a menace to social stability. By a curious anomaly a movement which sprang from the recognition of individual worth has also been adopted in those countries in which the political theory definitely and explicitly denies the worth of the individual.¹ In France fees have been abolished in the public *lycées* and *collèges* so that the authorities may be in a position to get rid of the *non-valeurs* or inferior students; in Germany the enrollments in the secondary schools have been reduced indirectly by the limitation of access to the universities, leaving for solution the serious problem of what to do with the excess; in England the introduction of alternative courses in the secondary schools of the West Riding of Yorkshire resulted from a study of the number of failures in the examinations based on the traditional secondary school curriculum.

What is obvious already in these movements is the recognition that if secondary education is to be provided for all, then it cannot be the same for all. And even in the traditional secondary school the relatively large number of failures raises three questions: How much real education have the failures received? How may courses be so diversified as to meet the needs of different types of abilities and aptitudes? and, How valid and reliable is the procedure of examinations? For the day seems to be passing when a high percentage of failures can be interpreted to mean the maintenance of high standards of achievement. Nothing is more certain than that much would be lost in lowering such standards; but it is equally certain that more can be done than has been done in the past (through improvement of instruction or through modified differentiation of curricula within the traditional concept of liberal education) to discover the type of education from which every pupil is capable of deriving some benefit,² and to set up a greater variety of standards suited to diversities of pupil abilities and aptitude.

If this presents an accurate picture of the situation in the secondary schools proper,³ the problem must inevitably become more complicated when

¹ And so Hitler can write: "The idea is intolerable that a hundred thousand completely incompetent men are considered worthy of higher education, while another hundred thousand of greater ability are left without any higher education. It is impossible to exaggerate the loss incurred thereby by the Nation."

"The national state must regard it as its most important task to open the doors of the public higher institutions to all talent, quite equally, no matter from what social group it comes. It must fulfill this task, for only in this way can the proper leadership of the nation arise outside of that group which represents a dead learning." *Mein Kampf*, 25th Edn., pp. 479 and 480.

² The psychological effects of success or failure upon the individual need not be discussed here.

³ See *A.M.A.*, Journal of the Assistant Masters' Association, January, 1927; County Council of the West Riding of Yorkshire, *22nd Annual Report of the Education Committee, 1925-26*, pp. 56 ff.; *The School Certificate Examination, being the Report of the Panel of Investigators appointed by the Secondary Schools Examinations Council to Enquire into the Eight Approved School Certificate Examina-*

16 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

some form of secondary education is open to all who have completed their primary education. The problem thus resolves itself into one of providing "the right education for the right pupil under the right teacher," to quote a definition of the function of educational administration by a distinguished English director of education, the late Sir Graham Balfour.¹

Leaving on one side for the present the question of developing the right methods for discovering the right education for the right pupil, it is obvious that any attempt to define "right education" at the level not only of secondary but of higher education as well must inevitably encounter the serious opposition of certain convictions and prejudices as to the meaning of culture or liberal education. The whole history of secondary education has, however, been the history of the conflict between traditional or established concepts of liberal education and attempts to redefine it in accordance with the demands of changing interests. The last three hundred years witnessed, first, the rise of demands for the inclusion of sciences, then of modern languages, and later still of other modern subjects in the curriculum of the secondary school. The process of adjustment was slow even though secondary education was still designed for a selected group.² It cannot be claimed that any definition of liberal education can be accepted as final, particularly at a time when it is proposed to provide some form of education for all at the secondary level. The problem has been beautifully defined by M. Léon Brunschvieg:

It is important that all the children of France should be considered alike as living plants, whose spontaneous growth will be assured by the same methods; only the trunk will be allowed to grow up to a certain height before the branches will be permitted to shoot out without the opposition of any artificial obstacle to the expansion of their being, whose innate powers will raise each up to the level designed for it.³

The same idea is inherent in a statement in the English Consultative Committee's Report on *The Education of the Adolescent*:

A humane or liberal education is not one given through books alone, but one which brings children into contact with the larger interests of mankind; and the aim of the schools in categories (II), (III), and (IV) above should be to provide such an education by means of a curriculum containing large opportunities for practical work and related to living interests.

tions held in the Summer of 1931 (London, 1932); W. Hartnacke, and E. Wohlfahrt, *Geist und Torheit auf Primarbanken* (Dresden, 1934); and *Educational Yearbook, 1930* (New York, 1931).

¹ Sir Graham Balfour, *Educational Administration*, p. 38 (London, 1921).

² See I. L. Kandel, *History of Secondary Education*, *passim*. (Boston, 1930).

³ Léon Brunschvieg, *Un Ministère de l'Éducation Nationale*, p. 79 (Paris, 1922).

In the earlier years the curriculum in these schools should have much in common with that provided in the schools at present commonly known as "secondary"; it should include a foreign language, subject to permission being given to omit it in special circumstances; and it should be given a "practical" bias only in the last two years.¹

There seems to be a widespread but unwarranted fear that any departure from the traditional concept of culture or liberal education will result either in lowering standards or destroying them altogether. Rather may it be said that any continued attempt to mould all pupils, irrespective of their interests and abilities, to the traditional pattern would lead to this result. The evidences of educational maladjustment, which the reports of the committees of the International Examinations Inquiry have already produced, if they have any meaning at all, point to the fact that *Non cuivis homini contingit adire Corinthum*, the haven of a liberal education, of *culture générale*, of scholarship as planned at present. It is too often forgotten by the staunchest and sincerest supporters of the traditional concept that they are the *beati possidentes* whose faith is strengthened by their own success and profit and who are prompted by that faith to employ the traditional pattern for the selection of an *élite* without regard for the large numbers who fall by the way.

It is obvious that it would be disastrous if standards of scholarship were to be sacrificed in order to provide a secondary education for all, but it would be equally disastrous if education becomes a mere competitive scramble for jobs. Nothing is so certain to-day as that the new age upon which the world has already entered demands nothing more than the widest dissemination of education; but equally nothing is more obvious from the evidence now available than that the task of educators is to redefine the concept of culture and liberal education in terms more suited to the world in which we live. Such a concept of culture, while holding fast to the best in the tradition, must be so flexible that it will meet the highly differentiated interests and abilities of the enlarged body of the adolescent group for which new educational opportunities are being planned. There are not many yet, in Europe at any rate, who would be ready to accept the statement of the problem as outlined by Professor C. Delisle Burns at the first Conference on Examinations:

I should like to suggest a more general reason for an enquiry into examinations than M. Bouglé has just suggested. It is this: that we in Western civilization are going through a transformation as fundamental, certainly,

¹ Board of Education, *Report of the Consultative Committee on The Education of the Adolescent*, pp. 84 f (London, 1926) The Roman numerals in the quotation refer to alternative types of schools at the secondary level.

18 PROBLEM OF EXAMINATIONS AND ITS SOCIAL SETTING

as the industrial revolution, and perhaps even as fundamental as the Reformation and the Renaissance. At present the type of person that we are producing for certain very important functions is not adequate to their fulfilment, and all the tests that we apply are antiquated. I am afraid that I should not agree with M. Bouglé, because he seems to imply in his assumptions that if the successful bank clerk and the successful industrialist were more like professors, they would be better persons in those positions. I very much doubt that. One of the worst troubles in the whole examination system is that it has been devised by professors, and the best thing that professors can think of is themselves; they therefore test candidates by what are tests of competence for professors, but not for bankers and other persons.¹

If the analysis of the present situation which confronts the educational world is accurate, it may be claimed that the reorganization which has taken place in secondary and higher education in the United States during the last thirty years represents an attempt to discover a solution of the type suggested by Professor Burns. This does not mean that the American solution has attained perfection. It is still open to many fundamental criticisms, that teachers are not adequately prepared for the new task; that there is an absence of thoroughness in the attainment of objectives; that no adequately recognized standards exist; that the system of credits, units or points has militated against the development of a sound education and established a largely irrelevant quantitative instead of a relevant qualitative measure of education; and that many of the changes have been forced on the schools by the tax-paying public or have resulted from too ready a desire to smooth the path for individual pupils and students. The period of reorganization has in any case been all too brief and the influx of students into the high schools and colleges has been too rapid to do more than make empirical adjustments to the

¹ *Conference on Examinations*, p. 226 (New York, 1931). See also the chapter of Professor Burns on Examinations and the Social Needs of the Modern World in *Essays on Examinations*, prepared by the English Committee of the International Examinations Inquiry, pp. 79 ff. (London, 1935), and the replies of laymen to a series of questions on the *baccalauréat* in the French Committee's report, *La Correction des Epreuves Ecrites dans les Examens*, pp. 14 ff. (Paris, 1936)

The same idea is conveyed in the following quotation from W. K. Hancock's *Australia*, p. 142 (New York, 1931) "Democratic sentiment applauds the sound argument that every office boy should have a chance to become a manager, and perverts it into a rule that no one shall become a manager who has not been an office boy. Australian Governments insist generally upon the rule that everybody must enter the public service at the age of sixteen or thereabouts. At the same time, by means of an excellent system of scholarships, they cunningly entice the cleverest boys to the Universities. When they have been enticed thither, these boys discover (unless they have entered upon a strictly technical training) that there is nothing for them to do except teach. So they return to school and encourage other clever boys to win scholarships. In this way the State has most ingeniously contrived that its system of democratic education shall not embarrass the public services by introducing into them resplendent talents. There has it is true been considerable reform in recent years."

demands of the situation. Professor Thomas H. Briggs, a recognized authority in the field of secondary education, has, indeed, made the statement that the United States does not have a philosophy of secondary education.¹ While the primary concern in the past two decades has centered upon the rapidly increasing body of students in secondary and higher education—the question of dealing with large numbers satisfactorily—there has been a growing concern in the last few years about the quality of education.

THE INTERNATIONAL EXAMINATIONS INQUIRY

It is at this point that the type of intellectual coöperation undertaken by the International Examinations Inquiry derives its meaning and significance. The European countries, it may be said, have stressed the maintenance of standards of quality in education with a consequent sacrifice of those whose abilities and aptitudes do not fit them into the traditional mould; the United States has stressed equality of educational opportunities, the result of which has been to open the high schools and colleges in practice to every boy or girl who seeks admission with a consequent sacrifice of quality. The European countries are ready to open the doors of secondary education more widely than hitherto; they may have much to learn from American practice, positively and negatively. On the other hand, the European countries are firmly resolved not to sacrifice the traditional standards of quality; in the present stage of unrest the United States may have much to learn from the further developments in Europe on this aspect of the problem. On both sides the fundamental issue is slowly ceasing to be one of selection and is becoming one of distribution, or the discovery of the right education for the right pupil under the right teacher.

Ultimately, then, the problem resolves itself into devising methods by which the right education can be discovered for the right pupil. The traditional method of selection has been by examinations. European studies of examinations up to the present seem to indicate that neither examinations as now conducted nor the teachers' judgments furnish reliable bases for such distribution.² It is on this question that far more experimentation and some progress, perhaps not always commensurate with the amount of experimen-

¹ See T. H. Briggs, *The Great Investment* (Cambridge, Mass., 1930) and *Secondary Education* (New York, 1933).

² Reference need be made here only to C. W. Valentine, *The Reliability of Examinations* (London, 1933); Sir Philip Harlog and E. C. Rhodes, *An Examination of Examinations* (London, 1935); O. Bobertag, *Schulerauslese* (Berlin, 1934); W. Hartnacke and E. Wohlfahrt, *Geist und Torheit auf Primanerbänken* (Dresden, 1934), and Commission Française pour l'Enquête Carnegie, *La Correction des Épreuves Ecrites dans les Examens* (Paris, 1936).

tation, have been made in the United States. The experiments and investigations began almost at the beginning of the present century. They opened with a study of the reliability of examinations, were followed by studies of differences of individual abilities and aptitudes, and were strengthened by the development of scientifically acceptable tests for measuring intellectual capacity and aptitudes. For educational purposes it was found that the derivation of intelligence quotients is too deterministic and that something more is needed to discover interests and aptitudes of pupils. The next step was the development of achievement tests and tests of aptitudes and to employ the results of such tests for the purpose of guiding pupils into those types of education from which they could derive the greatest profit. It was soon recognized that reliance could not be placed on one test alone or on a number of tests given at one time, and that if the best is to be done for each pupil the adviser or counselor must have the fullest information available about an individual both in and out of school. The result has been the development of cumulative record cards of such a type as to facilitate educational guidance. This development transcends the mere question of techniques of measurement and places in the forefront of educational statesmanship the task of discovering the right education for the right pupil. It is not intended to suggest that the final answer to this problem has been found in the United States; the question of educational values is not yet settled; society still has a stake in the kind of education which an individual receives, and the individual cannot be the sole measure of the education best suited to him.

The differences of background underlying educational systems, which were discussed earlier, have already indicated that the practices of one country may not be applicable to another. This much, however, is clear, that there is widespread concern everywhere about examinations, and that the results of the investigations conducted in those countries which are coöperating in the International Examinations Inquiry are in general similar to those reached in the United States some twenty-five years ago. The progress made since the establishment of the unreliability of examinations has been facilitated by the application of objective, statistical measurement to educational problems. Some of the experiments which have been conducted may have been crude; some of the expectations which flowed from the faith in the application of "scientific" measurement to the human being may have been extravagant; some of the tests which were invented may have been developed by the technician without an adequate mastery of the subject matter which he was trying to measure. None the less it must be admitted that many of the defects of the early experiments have been corrected and much of the skepti-

cism and resistance which the movement inevitably encountered at the start has been overcome. If the results of the investigations in the European countries are to bear fruit and if these countries are confronted with the crucial problem of the distribution and differentiation of education, much in American practice may be found suggestive for the next steps to be undertaken. The use of achievement and of placement tests, the results of the Pennsylvania Study conducted by the Carnegie Foundation, the employment of aptitude tests, and the development of cumulative record cards as instruments for educational guidance will have their contributions to make to European thought. On the other hand, it may be hoped that the projection of these techniques against the European background of culture and thoroughness may produce results which will be of significant value for American educators.

CHAPTER II

THE TRADITIONAL EXAMINATION

EARLY HISTORY

WRITTEN examinations appear to have been unknown in the United States until about the middle of the nineteenth century. Until that time schools and even colleges had annual "examinations" or inspections by school "visitors" or trustees. In Massachusetts an Act passed in 1789 made it the duty of the minister or ministers of the gospel and the Selectmen or others specially chosen by the locality to encourage school attendance

and once in every six months at least, as much oftener as they shall determine it necessary to visit and inspect the several schools in their respective towns or districts, and shall inquire into the regulation and discipline thereof, and the proficiency of the scholars therein, giving reasonable notice of the time of their visitation.¹

This provision only confirmed a practice which had already been introduced in Boston in 1709 when the town proceeded to

nominate and appoint a certain number of Gentlemen of Liberal Education, Together with some of the Rev^d Ministers of the Town . . . to Visit ye school from time to time, when and as Oft, as they Shall think fit, to Enform themselves of the Methodes Used in Teaching of the Schollars and to Inquire of their Proficiency . . . the Master being before Notified of their Comeing.²

In 1789 Boston appointed a visiting committee to carry out the requirements of the Massachusetts Act of that year. In addition to general inspection and supervision it was the duty of the committee

to examine the Scholars in the particular branches which they are taught, and by all proper Methods to excite in them a Laudable ambition to excel in a virtuous, amiable deportment and in every branch of useful knowledge.³

Such examinations were, of course, oral, and were conducted either by the teachers themselves under the supervision of the visitors or by the visitors themselves. Mass examinations could in fact not be conducted, since instruction was in the main individual, each pupil recited to his teacher, and advancement or promotion from one stage to another depended upon the teach-

¹ Quoted in Suzzallo, II, *The Rise of Local Supervision in Massachusetts*, p. 40 (New York, 1906).

² *Ibid.*, pp. 87 f.

³ *Ibid.*, p. 106.

er's opinion of the pupil's readiness. The earliest entrance requirements to Harvard simply provided that

When any Scholar is able to understand *Tully*, or such like classical Latine Author extempore, and make and speake true Latine in Verse and Prose, *suo ut aiunt Marte*; And decline perfectly the Paradigms of *Nounes*, and *Verbes* in the Greek tongue; let him then and not before be capable of admission into the Colledge.

It was not until 1734 at Harvard College and 1745 at Yale that examination by the president and tutors was added specifically to the requirements for entrance.

The external examination by a committee of visitors was the practice up to the close of the eighteenth century. Thus the Harvard College laws were revised in 1790 and

Among other improvements, "to excite the students to a noble emulation," a public annual examination was established, to be held in the presence of a joint committee of the Corporation and Overseers. This Committee, after examining the students of each class in the several branches studied by them during the year, were, with the President, Professors, and Tutors, to determine by ballot those who had distinguished themselves, and report their names to both boards, to be entered on record as a testimonial to their literary merit.¹

The examination was not popular and after some disturbances and petitions by the students the laws were again amended in 1796 and the control of the examination was delegated to the professors and tutors alone without the assistance of an external committee.

The general practice, then, both for entrance to college and in the college itself appears to have been the admission or promotion of students when they were "judged ripe." Provision was made for oral examinations but how rigorously this provision was enforced is not clear.²

In the secondary schools, the academies and later the high schools, examinations were public and were conducted quarterly or annually by committees. Thus in 1770 the Trustees of the Philadelphia Academy "considering that nothing can contribute more to keep up a laudable Emulation and Diligence among the Youth educated in this Seminary, than regular and stated exam-

¹ J. Quincy, *The History of Harvard University*, Vol. II, p. 279 (Boston, 1860). See also M. L. Smallwood, *An Historical Study of Examinations and Grading Systems in Early American Universities*, Ch. II (Cambridge, Mass., 1935), and S. E. Morison, *Harvard College in the Seventeenth Century*, pp. 67 ff., 81, 147, 206 f., 453, 458 (Cambridge, Mass., 1936).

² See E. C. Broome, *A Historical and Critical Discussion of College Admission Requirements*, p. 37 (New York, 1902).

inations of the different schools and bestowing Premiums and other marks of approbation on the most deserving Scholars" appointed four days in the year for such examinations.¹ There were no fixed practices. Examinations were held every quarter or every term (Leicester, 1821, and Lowell, 1832), or annually as well as at other fixed dates (Salem, 1836), or semi-annually (Salem, 1843), or even monthly (Portsmouth, 1847).² In 1845 Boston had two Sub-Committees, one of the Latin School and the other of the English High School. The first of these reported on August 5, 1845, "that they had examined said School on the 24th ultimo, and found it in its usual good condition." The second Sub-Committee after visiting the English High School on four different days in July for the purpose of making the Quarterly and Annual Examinations "were impressed with the thoroughness and accuracy manifested by the scholars generally in their various studies"; the proficiency in algebra and geometry was great but in French they noted a deficiency in pronunciation, but

The result of the examination has satisfied the Committee that the School maintains the high character which it has long held; and although they are far from intimating that there is no further room for improvement, they are fully of the opinion that this Board and the Public have no reason to be dissatisfied with its present condition.³

By the middle of the nineteenth century public examinations were generally held once a year and were more in the nature of public displays or exhibitions to show off brilliant pupils or to glorify the teachers. It was as a result of abuses to which such displays gave rise and of the criticisms which they prompted that written examinations began to be introduced and the public examination was replaced by graduation exercises and the distribution of diplomas to those who had passed a satisfactory examination conducted by the teachers.⁴

In the elementary schools the conduct of annual examinations by specially appointed committees was also the practice, but with the rapid increase in enrollments the difficulties and ineffectiveness of the system were recognized earlier at this educational level than in the secondary schools with relatively

¹ J. Mulhern, *A History of Secondary Education in Pennsylvania*, pp. 209 f. (Philadelphia, 1933).

² See E. D. Gizzell, *Origin and Development of the High School in New England before 1865*, pp. 331 f. (New York, 1923).

³ Quoted in O. W. Caldwell, and S. A. Courtis, *Then & Now in Education, 1845-1923*, p. 166 (New York, 1924).

⁴ See Gizzell, *op. cit.*, p. 334, and Mulhern, *op. cit.*, p. 140. Public examinations by the school boards or by sub-committees continued to be conducted in the 'sixties at the end either of each term or of each year; see H. Barnard, *American Journal of Education*, Vol. XIX, pp. 440 f.

smaller numbers. Thus in Boston, where two Sub-Committees were appointed annually, one to examine the Writing Department and the other the Grammar Department, the Sub-Committee for the latter reported in 1845 that it was "impossible to do anything like justice on the examination, for Committee-men could not be found who could give the time to examine over 7,000 children." Nor was the task less formidable when the examination was limited to the first class in each of the nineteen schools of the city. The Sub-Committee accordingly decided to introduce a written examination "in addition to the usual mode of oral examination." The printed questions for the written test were so graduated that the simpler could be introduced and answered by any children of common sense and ordinary attainments, while the more difficult would tax the powers of the best of our scholars without being quite beyond them.¹

The purposes which inspired the Committee in introducing what was regarded as an innovation in school practice are as interesting as they are modern in their definition of what should constitute a good examination.

It was our wish to have as fair an examination as possible; to give the same advantages to all, to prevent leading questions; to carry away, not loose notes, or vague remembrances of the examination, but positive information, in black and white; to ascertain with certainty what the scholars did not know, as well as what they did know; to test their readiness at expressing their ideas on paper; to have positive and undeniable evidence of their ability or inability to construct sentences grammatically, to punctuate them, and to spell the words.²

The Committee was desirous of having not only an adequate examination for each school, but a common examination which would yield results to make comparisons between all the schools of Boston possible. Accordingly the Committee asserted that

to the best of our knowledge and belief, the statistical returns which we submit will present the fullest and fairest means for judging of the real merits and demerits of our Schools, and of their comparative rank, which have ever been embodied so as to be within the reach of any but actual examiners of the Schools. We may say this safely, because it is perhaps the only statistical information which has ever been so embodied.³

Horace Mann commented enthusiastically on the written examination in

¹ See Caldwell and Curtis, *op. cit.*, p. 168. The full report of the Committee, as well as Horace Mann's comments on this innovation in *The Common School Journal*, Vol. VII, are reprinted in this book.

² Caldwell and Curtis, *loc. cit.*

³ Caldwell and Curtis, *op. cit.*, p. 170

The Common School Journal, Vol. VII. His enthusiasm was shared neither by the public nor by the teachers of Boston, particularly as the Committee had reported that "the first feeling occasioned by looking over these returns is that of entire incredulity" that so many pupils were unable to answer the questions set in the examination. Horace Mann was more farsighted than his teachers and analyzed the merits of the written examination with great penetration. (1) This method of examination is impartial, since the same questions are set to all pupils in the same class in all schools.¹ "Scholars in the same school, therefore, can be equitably compared with each other; and all the different schools are subjected to measurement by the same standard." Further, the questions in a written as contrasted with those in an oral examination are equal in ease or difficulty. (2) The new method is far more just than any other to the pupils themselves. In an oral examination of a whole class each pupil is questioned for at most two minutes; while in the written test he has a whole hour in which to arrange his ideas. (3) Accordingly the method under consideration is the most thorough, since pupils are not subjected to the chance of the few questions that can be given in the brief time of an oral examination but have a wider range suited to a greater range of attainment and ability. (4) The written examination does not, like the oral examination, give the teacher an opportunity to interrupt the procedure or offer suggestions to the pupils. (5) It removes all possibility of favoritism. (6) It determines, beyond appeal or gainsaying, whether pupils have been faithfully and competently taught, for while the oral question tends to call forth a factual answer, in the written examination the pupils are able to develop ideas and show the connections of facts. (7) Finally, in a written examination "a transcript, a sort of Daguerreotype likeness, as it were, of the state and condition of the pupils' minds, is taken and carried away, for general inspection"; that is, a permanent record is available by which schools may be compared with each other or each school may measure its own progress. A recognized standard of comparison is thus available, as contrasted with the different standards of judging in the minds of different men, for "if every man's foot is to be taken as twelve inches long, it becomes an important question by whose foot we shall measure."² Horace Mann thus not only defined the fundamental problems in any system of examinations, but anticipated by implication the need of some form of recognized standardization.

¹ It is curious that the Sub-Committee had not hit upon the idea of setting the same examination at the same time, but went from school to school as rapidly as possible in order to prevent any leakage of the questions.

² Caldwell and Courtis, *op. cit.*, pp. 238 ff.

The superintendent of the Chicago schools reported in 1857 the introduction of a written examination for admission to the public high school. "Every effort is made to put the candidates as much at their ease as possible, and secure them from all unnecessary embarrassments." To prevent favoritism or bias the candidates were given numbers, their names being recorded on identification slips. "Besides the teachers of the school, on whom the examination chiefly devolves, several members of the Board of Inspectors and the superintendent are in constant attendance, aiding and directing in the different exercises." Great care was taken to secure reliability and uniformity in the standard of judgment; the papers whose averages came within five or ten per cent of the lowest rank admitted were rechecked by the principal of the school and the superintendent. The written examination, it was claimed, gave all candidates a better opportunity of doing themselves justice than did an oral examination. Of all experiments with examinations "the written examinations afford the most reliable test of qualifications, and are on the whole the most just and satisfactory to all parties. This mode of examining is now adopted in nearly all the principal cities of the Union."

The method of the written examination was adopted in 1856 not only for admission to the public high school but for purposes of promotion. In the opinion of the superintendent

These written reviews are among the most successful means for securing thoroughness and accuracy of scholarship. . . . This mode of examining a class accomplishes at least three important objects at the same time. It affords a thorough test of the pupil's knowledge of the subject; it is one of the best methods of cultivating freedom and accuracy in the use of language; and it furnishes a valuable discipline to the pupil's mind, by throwing him entirely on his own resources.¹

EARLY CRITICISMS OF EXAMINATIONS

Opposition to examinations began in the eighties. They were criticized because of the stress on uniformity, "a system of strait jackets" as they were called by one superintendent, and because they encouraged pupils to study for marks. In the words of Superintendent Streeter of Titusville, Pa.,

There is nothing definite about them. No two teachers would mark the same paper alike; and I have sometimes wondered how much a night's rest

¹ H. Barnard, *American Journal of Education*, Vol III, pp. 532 ff, gives a portion of the *Third Annual Report*, 1856, of W. H. Wells, Superintendent of Schools in Chicago

or a breakfast has to do with papers marked 98 and 99 on the one hand, and 15 and 0 on the other.¹

There began to develop at this time a lack of confidence in the value of marking and in the justice of examination systems. And yet there was recognized the need of some method of standardizing the work of the schools and of some measures by which the work of schools over a large area could be compared. Thus Superintendent B. F. Patterson of Pottsville, Pa., in his report to the State Superintendent in 1888 writes

We have always thought that there should be some means by which we would be enabled to compare our work one with another. . . . What we would like to have is some test by which we can measure work. . . . Let us stop talking and compare our work. . . . But we would like to have comparative statistics on other subjects, for instance: How long should it take the average pupil, in reading, to reach the fourth reader? Again, what per cent of our monthly enrolment are in fourth readers or beyond it? At what average age should pupils be expected to readily solve miscellaneous problems in the fundamental rules; such problems as they will meet in any ordinary calling. Each borough superintendent has the means of knowing the comparative standing of each school of the same grade in his own district; but we are all unable to compare our own schools with those of any other district in the State.²

The enthusiasm with which Horace Mann welcomed the introduction of the written examinations began to wane before the close of the century. At the high school level the domination of college entrance examinations began to be attacked. In the practice of the elementary schools examinations for promotion were subjected to criticism. In both cases the burden of the criticisms were based on the recognition that "uniformity of conditions in pupils does not exist"; pupils are unequal in mental ability, vigor, ambition, application, home advantages, and health. They begin the school year with unequal attainments and finish with greater differences because of a lockstep system, crowned by examinations applied equally to the able, the average, and the poor pupils.

The use of stated examinations, wrote Superintendent E. E. White of Cin-

¹ Quoted from the *Pennsylvania School Journal*, Vol. 35, 1886, p. 128, in J. Mulhern, *A History of Secondary Education in Pennsylvania*, p. 593. The same idea was given poetical form by A. T. Godley in his "Rubaiyatt of Moderations" (in *Lyra Frivola*, Methuen, London, 1899):

'Twixt Right and Wrong the Difference is dim;
'Tis settled by the Moderator's Whim:
Perchance the Delta on your Paper marked
Means that his Lunch has disagreed with him.

² Quoted in Mulhern, *op. cit.*, p. 594, from the *Report of the State Superintendent*, 1888, pp. 163 f.

cinnati, as the sole means of determining the fitness of pupils for promotion has been attended with serious perversions of true school work. It has narrowed and grooved instruction, encouraged the use of mechanical and rote methods, and occasioned cramming and vicious habits of study.¹ As a consequence faith in the value of the written examination as a method of determining fitness for promotion had been shaken.

A few years ago the use of the written examination as a basis of promotion was well-nigh universal in graded schools. But the serious evils attending the examination system have demanded wide attention, and, as a result, many cities have either modified the system or abandoned it altogether. The evidence of these changes is as conclusive as it is gratifying.²

The result of the unrest was indicated in the answers to a questionnaire which showed that, while the written examination was still used to determine fitness for promotion from class to class and from school to school, there was a strong tendency to rely on the judgment or estimates of the teachers, sometimes alone, sometimes supplemented by an examination. All the arguments pointed to the elimination of stated examinations, which were practically external in character. It was found that there was great variation in the marking, ranging from one to twenty per cent in different schools. But what was regarded at that time as a more serious indictment of the system of examinations was their effect upon instruction and methods of study. "The guiding question becomes not what is best for the pupil, but what will 'count' in the examination." It was recognized that "the best study is done where there is the freest play of natural motives, and the poorest study where there is the most absorbing interest in examination marks," leading to overpressure, strain, waste, dishonesty, and miseducation.

For the stated examination were substituted frequent examinations by the teachers for instruction purposes, and cumulative records of the pupils' standing. The leading cities were beginning to promote on the judgments of the teachers and on the class standing of the pupils without stated examinations.

We thus reach the decisive question involved in the promotion of pupils on their success in daily work. This success must be reliably determined, and the only means of doing this, in the absence of examinations held for the purpose, is the teacher's judgment, recorded or otherwise, and supported, as it may usually be, by the judgment of the principal.³

¹ E. E. White, *Promotions and Examinations in Graded Schools*. U. S. Bureau of Education, Circular of information, No. 7, 1891, p. 14

² *Ibid.*, p. 28.

³ *Ibid.*, p. 43.

Teachers, it was argued, are as competent to determine the comparative success of their pupils as they are to mark examination papers. In Cincinnati examinations for promotion had been abolished in 1886 and replaced by the teachers' estimates. In the first year of the transition both examinations and estimates were used, but at the close of this year there was a general concurrence in the belief that teachers' estimates in the last five months represented the fitness of pupils for promotion more reliably than the results of examinations in the first five months of the year. Fewer pupils were promoted on the estimate plan than on examinations; and although pupils who were not promoted could claim the right of a written examination, the number of such applications decreased each year. The number of applicants who succeeded in reaching the standards for promotion on the basis of an examination was almost negligible, thus corroborating the reliability of the teachers' judgments. The teachers recorded careful estimates of the pupils' fidelity and success in the different branches of instruction without daily marking or examinations. The inquiry concluded, however, that the continuance of examinations by the teachers was desirable for instruction purposes. The teacher "must know to what extent his instruction has been appropriated by each pupil in his class," and "testing is as essential a part of teaching as instruction itself," and may lead to earnest re-teaching of a subject.

At this time President Charles W. Eliot of Harvard College began his attack on examinations and his plea for an adequate consideration of individual differences. In his inaugural address, delivered in 1869, he had not only insisted upon strict standards of admission to prevent "a waste of instruction upon incompetent persons," and annual examinations, but had deplored the absence in American colleges of external examinations of the type found in Europe.

When the teacher examines his class, there is no effective examination of the teacher. . . . The American practice of allowing the teaching body to examine for degrees has been partly dictated by the scarcity of men outside the faculties who are at once thoroughly acquainted with the subjects of examination, and sufficiently versed in teaching to know what may fairly be expected of both students and instructors. This difficulty could now be overcome.¹

In a discussion of *The Gap between Common Schools and Colleges* (1890) President Eliot urged the careful preparation and marking of examinations, but on the whole seemed inclined to favor the further extension of systems of inspection.

¹ C. W. Eliot, *Educational Reform*, pp. 9 f. (New York, 1901).

We may expect state examining and inspecting systems to improve and extend, for they have demonstrated their utility; and, remembering the extremes to which examination methods have been carried in England, we may reasonably hope that state boards will more and more inspect institutions as well as examine individuals. The profession of school inspector will, therefore, become recognized as a separate and honorable calling.¹

Two years later President Eliot fell in with the prevailing criticisms of examinations but approached the problem from an entirely different angle, anticipating later thought on the subject by nearly a quarter of a century. His attack on mass examinations was stimulated by his fear of uniformity, of the lockstep in education, of the attempt to treat all pupils and students alike. Other things being considered he wished the theory upon which he had introduced electives at Harvard College to be applied at all levels of education. This meant, of course, a more rational interpretation of the application of the democratic theory to education than was popular either in his day or even in our own time. Thus in his address delivered before the National Education Association on *Shortening and Enriching the Grammar School Course* he defined the problem as follows:

The democratic theory, it is said, implies equality among the children, uniformity of programme, uniform tests for promotion, and no divisions in the same schoolroom according to capacity or merit. I need not say to this audience that these conceptions of true democracy in schools are fallacious and ruinous. Democratic society does not undertake to fly in the face of nature by asserting that all children are equal in capacity or that all children are alike and should be treated alike. Everybody knows that children are infinitely diverse—that children in the same family, even, are apt to be very different in disposition, temperament, and mental power. Every child is a unique personality. It follows, of course, that uniform programmes and uniform methods of instruction, applied simultaneously to large numbers of children, must be unwise and injurious—an evil always to be struggled against and reformed, so far as the material resources of democratic society will permit. It is for the interest of society as well as of the individual, that every individual child's peculiar gifts and powers should be developed and trained to the highest degree. Hence, in the public schools of a democracy the aim should be to give the utmost possible amount of individual instruction, to grade according to capacity just as far as the number of teachers and their strength and skill will permit, and to promote pupils, not by battalions, but in the most irregular and individual way possible. . . . Uniformity is the curse of American schools. That any school or college has a

¹ *Ibid.*, p. 218.

uniform product should be regarded as a demonstration of inferiority—of incapacity to meet the legitimate demands of a social order whose fundamental principle is that every career should be open to talent. Selection of studies for the individual, instruction addressed to the individual, irregular promotion, grading by natural capacity and rapidity of attainment, and diversity of product as regards age and acquisitions, must come to characterize the American public school, if it is to answer the purposes of a democratic society.¹

The same thesis was repeated in the same year (1892) in another address on *Undesirable and Desirable Uniformity in Schools*, an address in which the place of examinations in the interpretation of democratic education was defined.

Flexible and diversified school programmes will give all the children their most favorable chance; stiff and uniform programmes will not. . . . There must be examinations, anticipated and unanticipated. Let them always be conducted by the teacher, for the teacher, and as helps and guides in teaching and learning. The teacher needs to ascertain, from time to time, by such tests how his instruction has been assimilated by his class; and the pupils need to learn from them what the teacher expects of his class.²

President Eliot continued the same plea for adaptation to individual capacities in 1897 in a discussion of *The Function of Education in Democratic Society* in terms which constitute one of the main theses of the Pennsylvania Study conducted by the Carnegie Foundation (see pp. 133 f.).

Another important function of the public school in a democracy is the discovery and development of the gift or capacity of each individual child. This discovery should be made at the earliest practicable age, and, once made, should always influence, and sometimes determine, the education of the individual. It is for the interest of society to make the most of every useful gift or faculty which any member may fortunately possess; and it is one of the main advantages of fluent and mobile democratic society that it is more likely than any other society to secure the fruition of individual capacities. To make the most of any individual's peculiar power, it is important to discover it early, and then train it continuously and assiduously. It is wonderful what apparently small personal gifts may become the means of conspicuous service or achievement, if only they get discovered, trained, and applied. . . . The vague desire for equality in a democracy has worked great mischief in democratic schools. There is no such thing as equality of gifts, or powers, or faculties, among either children or adults. On the con-

¹ *Ibid.*, pp. 262 and 265.

² *Ibid.*, pp. 281 and 285.

trary, there is the utmost diversity; and education and all the experience of life increase these diversities, because school, and the earning of a livelihood, and the reaction of the individual upon his surroundings, all tend strongly to magnify innate diversities.¹

The theory of diversification and differentiation of education was sound; there were, however, no instruments or techniques available whereby individual differences of capacities and interests could be discovered or measured. Nor was the United States any more ready than the European countries at that time to redefine the concept of education and culture, although the unrest was just beginning in the last decade of the nineteenth century. President Eliot was sound in decrying uniformity of aims and standards, but he did not recognize or suggest any substitutes for examinations other than the teachers' judgments nor did he realize the difficulties involved in the conduct and marking of examinations.

The earliest attack on the second of these problems was made in a paper on *The Right Reform of Examinations*, read at the First Annual Convention of the College Association of the Middle States and Maryland (1890) by J. Rendell Harris, Professor at Haverford College. Professor Harris, while he acknowledged the values claimed for examinations, such as recapitulation and concentration of studies, suggested that "the first thing to be reformed is the examiner." Examinations would not be improved until the examiner had a clear notion of what to aim at. The attempt to make all of a group of students satisfy a certain task equally is a *reductio ad absurdum*; "methods which aim at democratic results have no place in examinations." Nor is the purpose of examinations one of passing some students and failing others but of discovering how the students stand in relation to each other. "A well conducted examination divides the students one from another like the opening out of a fan. I affirm that the first thing to be aimed at is to produce a dispersion among the group of persons presented for examination," and the more dispersion a teacher produces, the better examiner he is.

Our purpose is to show how unequal students are to one another; and the right way to do this is not by setting up a standard of passing or failing, as if there were only two conceivable students A and not A (the elect and the non-elect), nor by the *a priori* assumption that there are four conceivable classes of students, say A, B, C, D, of which D stands for the non-elect; but by recognizing that there are in reality as many classes as there are students, and trying to make this fact as patent as possible by the

¹ *Ibid.*, pp. 408 f.

process of examination, we can come to the question of ticketing or bracketing afterward.¹

This means that examiners and teachers when they profess to give certain grades, say on a scale of 100, are proceeding on an artificial theory for which there is no foundation, as is shown in the variation of marks for the same paper by different markers.

I repeat, however, that the whole process of classification of students is an induction, based upon the observation of dispersion. And it follows, therefore, that the democratic method, which aims at producing the greatest possible number of equal and similar molecules, each ticketed with the same alphabetical symbols, is unscientific, since it reduces dispersion almost to the lowest possible point.²

An individual's place on a normal curve of distribution, which is what Professor Harris means by "dispersion," is more important than an alphabetical or numerical mark. Professor Harris deplores the absence of standards (there are "as many currencies as there are institutions") resulting from the practice by which teachers are responsible for training, examining, and marking their own students without any external check. Like President Eliot he recommended the creation of some body of external examinations which would contribute to the improvement of secondary education "with the full intention that we will gather only the good fish into their appropriate vessels and throw the bad away."³

RISE OF THE ACCREDITING SYSTEM

Professor Harris was some fifteen years ahead of his time in his method of approach to examinations. The more technical study of the problem had to await the introduction, first, of statistical methods in educational research, and, secondly, of standardized norms in the form of scales and tests. The practice of conducting external examinations of pupils in elementary schools had, with few exceptions, disappeared before the close of the nineteenth cen-

¹ *Proceedings of the First Annual Convention of the College Association of the Middle States and Maryland, 1890*, p. 90.

² *Ibid.*, p. 94.

³ Professor Harris, a philologist, paleographer, and biblical scholar, had come from Cambridge University in 1882 to accept a professorship at Johns Hopkins University; in 1886 he joined the faculty of Haverford College. He admitted that his own standards had suffered because of the absence of external checks and his personal responsibility to his students. "In dealing with the students, my aim is to let them all through, or nearly all, on as good a grade as possible, because I am one of those persons who would not willingly set foot on a worm, especially worms that would turn—and rend you." (*Ibid.*, p. 96.)

It is to be noted that Professor F. Y. Edgeworth's attack on examinations appeared at this time. In

tury. The problem was most pressing at the point of relationship between high schools and colleges. This, indeed, was the chief preoccupation in the last decade of the nineteenth century. It was generally charged that the high schools could not carry out their aims and purposes as long as they were dominated by inelastic and inflexible requirements for admission to college and so long as these requirements differed for different colleges.

The problem had been settled to some extent in the Middle West and West by the adoption of the accrediting system which had been first adopted by the University of Michigan in 1870. The system, while it abolished entrance examinations and introduced the practice of inspecting high schools, did not at that time develop any greater degree of flexibility in the entrance requirements. It simply meant that students who had completed these requirements and were recommended for admission to the University by the principal of a school accredited after inspection could be admitted without examination. Changes and modifications in the entrance requirements were to follow much later as the enrollments in the high schools expanded and the number of students proceeding to higher education increased. More recently it has come to be realized that even a system of inspection and external standards for accrediting of schools fails to provide an adequate method of standardizing the quality of instruction and attainments. The somewhat vague and varied requirements for certifying high school teachers and the equally vague standards of marking among a large body of teachers has led to the adoption of academic tests, sometimes statewide, occasionally regional, in scope to supplement other methods of accrediting.¹

ADMISSION TO COLLEGE BY CERTIFICATION

The unrest caused by the so-called dominance of college entrance requirements and by the desire to allow greater flexibility to the high schools was to some extent met by the adoption of the accrediting system in the Middle West and West. In the New England States the same difficulty was met by the establishment in 1902 of The New England College Entrance Certificate

an article on "The Statistics of Examinations" (*Journal of the Royal Statistical Society*, Vol. 51, 1888, pp. 599 ff.) Edgeworth wrote "That the examination is a very rough yet not wholly inefficient test of merit is generally admitted. But I do not know that anyone has attempted to appreciate with any approach to precision the degree of accuracy or inaccuracy which is to be ascribed to the modern method of estimating proficiency by means of numerical marks. It occurs to me that the method of effecting this computation is contained in that part of the Calculus of Probabilities which is known as the Theory of Errors. The doctrine of chance and error is eminently germane to the present subject." This article was followed two years later by another on "The Element of Chance in Competitive Examinations" (*Journal of the Royal Statistical Society*, Vol. 53, 1890, pp. 460 ff. and 644 ff.).

¹ See pp. 104 f.

Board. The following institutions were represented on the Board: Amherst College, Boston University, Bowdoin College, Brown University, Dartmouth College, Mount Holyoke College, Smith College, Tufts College, University of Maine, University of Vermont, Wellesley College, Wesleyan University, and Williams College. The function of the Board was to draw up a list of approved secondary schools from which students would be admitted to one of the constituent colleges without examination on the certification of the principals.

To be placed upon the approved list schools were required to submit with their applications full information concerning their courses of study, teachers, and equipment. The preparation given in such schools had to follow one of the recognized plans for entering one of the colleges represented on the Board. Approval was granted to schools for periods of three years on the basis, first, of the records of their students who had already been admitted to the colleges, and, secondly, on the basis of the records of students admitted subsequent to the time of approval. As contrasted with the accrediting schemes the certification plan did not involve actual inspection of schools, but laid stress on the work of students from the approved schools during their first year in the colleges concerned.

ESTABLISHMENT OF EXAMINATION BOARDS

The issue was more acute in the Eastern states where each college had its own requirements for admission. Here the institutions of higher education were private and the same relations could not be established as in those states where both the universities and the high schools were provided and maintained by public authorities. Here the crucial issue was "to devise means for securing greater uniformity in college admission examinations." A concerted movement in this direction began in 1879 when an agreement was reached at a Conference of New England Colleges to accept the standards set by Harvard in the examination for English. This practice was extended to other subjects and in 1885 there was created the New England Association of Colleges and Preparatory Schools for "the advancement of the cause of liberal education by the promotion of interests common to colleges and preparatory schools." The Association of Colleges and Secondary Schools of the Middle States and Maryland, which had already begun to meet in 1887, was formally established to carry out similar purposes. Out of these demands and movements toward uniform college entrance examinations there was established the College Entrance Examination Board, thus putting into effect

a proposal made as early as 1877 and frequently repeated thereafter by President Eliot¹ and carried through to success by President (then Professor) Nicholas Murray Butler.

There was thus established an agency to conduct examinations in different parts of the country and abroad, and to award certificates to be accepted by college admission boards in lieu of their own separate admission examinations. The creation of the College Entrance Examination Board introduced no innovation in the conduct of examinations; it merely responded to one of the demands already expressed by the Committee on College Entrance Requirements "to study the question of college entrance examinations for the purpose of harmonizing the relations between the secondary schools and the colleges."² It was not the purpose of the Board to standardize secondary education throughout the country, but it did establish through its examinations standards and prescriptions for pupils who were planning to enter certain colleges. To a certain extent standardization of high schools was promoted in many states by the accrediting system, in certain groups of states by regional standardizing agencies, and for the country as a whole through the reports of the Committee of Ten of the National Education Association (1893) and of the Committee on College Entrance Requirements (1895), which helped to create a national currency of norms in secondary education.³

REGENTS' EXAMINATIONS IN NEW YORK STATE

The control of secondary education through examinations was nowhere as strict and direct as in New York State. Here the Board of Regents, established in 1784, began in 1865 to conduct preliminary examinations at the close of the elementary school course as a basis for the distribution of the Literature Fund. In 1878 the Board, in response to a request of the principals of academies undertook to examine secondary schools "for the purpose of measuring the work of the schools and establishing proper standards of scholarship" and to "furnish a suitable standard for graduation from academies and academic departments of union schools, and of the admission to the several colleges of the State," as well as for distribution of support from the Literature Fund.⁴ In 1900 the practice of paying grants on the basis of

¹ See C. W. Eliot, *Educational Reform*, p. 219 (New York, 1901).

² For the resolutions of the Committee on College Entrance Requirements see E. C. Broome, *A Historical and Critical Discussion of College Admission Requirements*, pp. 133 ff. (New York, 1902)

³ See I. L. Kandel, *History of Secondary Education*, pp. 469 ff.

⁴ The influence of the Regents' Examinations upon the work of the schools cannot be discussed here. An investigation by W. W. Knox, *A Study of Some of the Influences of Regents' Examinations in*

the examinations was changed and a return was made to payment on attendance and reports of inspectors. In 1906, however, the payment of grants was made dependent on the use of the Regents' examinations but not on the results. Beginning in 1880 the Board published syllabuses to furnish guidance to those schools which submitted pupils for the examinations. Discussing the practice in 1890 President Eliot admitted the "stimulating, elevating, and unifying influence of these examinations," but criticized both the preparation and the marking of the papers.¹

The examination papers were prepared by a staff of inspectors and examiners under the direction of the secretary of the Board until 1906, when the coöperation of the schools and teachers was secured in their preparation. From this time on papers in the major subjects have been prepared by committees consisting of representatives from high schools and colleges and the state supervisors of the respective subjects. The papers so prepared are submitted to critical examination for content and form by the editorial staff of the State Department of Education, and then resubmitted for final scrutiny to a Board of Revision, usually of seven members, appointed by the State Examinations Board, which was created by Dr. Andrew S. Draper, then State Commissioner of Education, and consists of twenty members—fifteen representing colleges, superintendents of schools, and high school teachers, and five representing the State Department. The Board of Regents, in shaping the policies of the State with reference to courses of study and examinations has thus endeavored to utilize the experience and judgment of those actually engaged in teaching the prospective candidates for the examinations.

In the preparation of the examinations the claim has been made since 1920 that care is taken to test power, mastery, and comprehension rather than knowledge or the acquisition of facts, which place a premium on application² and a good memory.

The older methods of teaching were concerned largely with memory processes and the resulting tests were largely fact questions seeking to find out what the pupil had remembered of a given mass of information. Newer methods of teaching seek to train reason and judgment, are concerned with cause and effect, with the marshaling of evidence from which the pupil is to

Chemistry, concluded that the effect of the system is to lead teachers to emphasize topics in chemistry stressed in past examinations, to plan courses in accordance with the syllabus requirements, to give special attention to inferior pupils who were likely to fail in the examinations, and to introduce new topics only as they were required by the Regents (Unpublished dissertation, Teachers College, Columbia University, 1933)

¹ C. W. Eliot, *Educational Reform*, pp. 207 ff (New York, 1901).

² To the extent that tests do measure application (industry) they would be desirable.

deduce conclusions, and seek to establish likenesses and differences, comparisons and contrasts. The resulting tests, therefore, are largely thought questions measuring the pupils' understanding and appreciation of the subject studied.¹

In 1923 after some years of discussion the Department began to experiment with the new, objective, short answer type of examination in English grammar and silent reading; in 1924 about half of the questions in algebra and geometry were of the new type, and true-false and completion tests were included in the social science paper. It was felt, however, that care must be taken that, besides measuring only factual memory, efforts must be made to retain tests of power. In 1925 an experiment was conducted, under the direction of Dr. Ben D. Wood, in modern languages and physics. The regular three-hour examination was divided equally between the old and new type questions, or, as they are called in the Department's reports, subjective and objective type questions. Dr. Wood's report on the experiment was that, in comparison with the old type examinations taken at the same time, the new type tests were much superior in objectivity, reliability, and validity, and that, while the cost of preparing and printing the new type tests was higher, the cost for marking or rating the papers was lower. Dr. W. W. Coxe of the Board's staff agreed with Dr. Wood on the general principles of the new type examinations but not on their superiority.²

Two reasons impelled the Board of Regents to adopt the new type examinations, even though only in part. The first was the criticism of the examinations and their results from the proponents of scientific measurement in education,³ and the second was the rapid increase in the number of papers to be marked without a corresponding increase in the number of examiners. By 1929 the Board had come to the conclusion that the ideal examination is a combination of the new and old type of questions, one to measure factual knowledge, and the other mastery or ability to organize a mass of material and to present it in narrative or expository form; "one is a test of cumulative knowledge, the other of assimilated knowledge."⁴ Objective examinations have been gradually extended year by year and are now found in whole or in

¹ University of the State of New York, *Bulletin to the Schools*, Vol. 12, October 1, 1925, p. 19. See also *Annual Reports of the Education Department*, 1923, p. 65, and 1924, p. 223.

² *Annual Report*, 1927, p. 47. Professor Ben D. Wood's report on the new type examinations in modern languages and physics was published under the title *New York Experiments with New-Type Modern Language Tests* (New York, 1927).

³ Joint Committee on Rural Schools. *Rural School Survey of New York State*, Vol. II. *Administration and Supervision*, Part V, pp. 414 ff. (Ithaca, N. Y., 1923).

⁴ See *Annual Report*, 1929, Vol. I, p. 188; and *Annual Report*, 1933, Vol. I, p. 212.

part in practically all the subjects offered, the proportion varying somewhat with the nature of the subject.

The examination papers are marked first by teachers or committees of teachers in the schools where they are written, so that the pupils might be said to have the advantage¹ of having their papers marked by those who are familiar with their work. Detailed suggestions and instructions for marking, prepared with their coöperation, are issued to teachers. The passing papers, or, as they are called, the papers "claimed," are forwarded to the State Department for rereading by the staff examiners assisted during the summer by more than one hundred and fifty experienced teachers who must be college graduates and must have had more than three years of experience in teaching the subject in which they examine. Papers from schools with outstanding records on examinations and in inspectors' reports are reviewed by sampling, and only when the sampling reveals a failure to maintain standards are all the papers reread. In cases where the marks are debatable the papers are given a further reading and no paper is rejected until it has had a second reading, usually by a senior examiner or the supervisor. "All scholarship and pre-professional papers are read twice by different examiners and in case of material variation are given a third reading." How formidable the task of reading is may be gathered from the increase in the number of papers which rose from 1,126 in 1878 to 330,226 in 1920 and 1,117,516 in 1934.

It is claimed for the Regents' examinations that the method of preparing the examination questions guarantees close adaptation² to the work actually done in the schools, and that this work is in turn guided by the syllabuses which are also the result of coöperation between officials of the State Department and the school teachers. A proposal made in 1923 to substitute a system of certification of pupils by the schools for the examinations was rejected on the ground that the examinations "afford a satisfactory means of maintaining proper scholastic standards, of improving the quality of teaching, and of providing a satisfactory method of admission to higher institutions."³

In comparing the predictive value of new and old methods of admission to college, Dr. Ben D. Wood found that of four methods (Thorndike Intelligence Scores, Regents' examinations, College Entrance Board, and secondary school marks) the Regents' examinations correlated nearly as high as the

¹ At the same time the disadvantage of having a great variety of different standards which may result from an absence of uniform qualifications for teachers must be recognized and admitted. The practice still leaves a large part of the responsibility for securing uniform standards on the central reviewing body, which may or may not be equal to the task.

² But not necessarily to the individual student who is able to go beyond the prescribed requirements.

³ University of the State of New York, *Bulletin to Schools*, Vol. 12, October 1, 1925, p. 19.

Thorndike tests with marks obtained by students in their freshman and sophomore years in college, and considerably higher than the certification method on the basis of secondary school marks. Two important differences are to be noted, however,—first that the Regents' examinations at the time of the comparison were still of the traditional, subjective type, while the Thorndike tests were of the new objective type, and, secondly, that the latter took only three hours, while the former required fifteen hours.¹

The Regents' examinations have always been the subject of criticism, at one time because of their effects upon the work of the school, but later and more persistently because of variations in results and consequent unreliability. The later criticism coincided with the development of statistical measurements in education. Thus the *Rural School Survey*, which was published in 1923, pointed out that there were large variations in the percentage of failures from year to year in a given subject and among different subjects in a given year; that there were discrepancies between the percentage of papers "claimed" as of passing grade by the teachers and the percentage "accepted" or finally passed by the examiners; that these discrepancies varied from year to year and from subject to subject; that the distribution of marks showed no consistency among themselves as to form and no consistent tendency to approximate the normal curve to be expected in the distribution of homogeneous groups; and, finally, that the ratings are not standardized or reliable.

Despite the systematic efforts, conducted over a period of twenty-five years, to secure cooperation between examiners and teachers, the variations in results and discrepancies in standards continued. Thus in elementary algebra the range of papers "accepted" by the examiners was from 50 per cent in 1897 to above 75 per cent in 1907, and approximately 60 per cent in 1920; in biology 85 per cent of the papers were "accepted" in 1907 and 52.5 per cent in 1920; in physics the range of "accepted" papers was from 35 per cent to about 80 per cent in this period; and in American history from 25 per cent in 1898 to 86.4 per cent in 1920. Within these ranges there were marked fluctuations from one year to the next; in American history 25 per cent of the papers were passed in 1898 and 61 per cent in the following year; in physics 63 per cent were passed in 1898 as compared with 36 per cent in

¹ Ben D. Wood, *Measurement in Higher Education*, pp. 86 ff (New York, 1923). On the other hand, Professor Paul Kruse after considering the same question, the predictive value of Regents' examinations, stated at about the same time that "It seems fair to conclude that the Regents' examinations serve no better than do teachers' marks as measures of fitness for college, and that they probably serve less well." See Joint Committee on Rural Schools, *Rural School Survey of New York State*, Vol. II, *Administration and Supervision*, pp. 456 ff. (Ithaca, N. Y., 1923).

1897. In the same way there were great variations in any one year in the percentage "accepted" in the different subjects of the examinations; the range in 1920 was from 52.5 per cent in biology to 86.4 per cent in American history. According to the *Rural School Survey of New York State*

We have here evidence that there is great variation either in the difficulty of the examinations from year to year and from subject to subject in terms of the percentage of pupils writing who achieve a passing grade, or great variation in the rating of the papers, or both.¹

If the following assumptions are accepted, that the ability of pupils and of the teachers in a given year will not vary greatly, and that the subject matter taught will not vary greatly from year to year, then it might be expected that the same percentage of pupils should achieve a passing grade in the various subjects of the school course, provided that the work is adjusted to their needs. The results up to the time of the survey (in 1922) seemed to show that chance played a large part in determining the percentage of pupils who passed in any subject in a given year.

Not only were the results as finally determined by the examiners variable but they disagreed with the standards of the teachers. The average difference between the percentage of papers "claimed" as of passing standard by the teachers and the percentage "accepted" by the examiners over a period of twenty-five years was 9.4 per cent, and in twelve subjects 9.2 per cent.²

The experiments with new type tests began after these critical observations of the examination results over a period of twenty-five years were made. New type tests have been introduced since that time in more subjects and in larger proportions in each subject. It is interesting accordingly to compare the results (on p. 43) over a period of five years preceding the inquiry (1916-1920) with the results in the last five years (1931-1935).

It is unfortunate that the results of the old and new type examinations are not separated in the reports; it would then be possible over a period of years to compare the reliability and validity of the two types. Since the general conditions have remained the same with one exception (the unparalleled increase in the number of pupils in the secondary schools), the relative stability of the results in the last quinquennium as compared with the period from 1916 to 1920 can hardly be attributed to improved techniques of marking the subjective type of the examinations. Whether the new type tests are

¹ Joint Committee on Rural Schools, *Rural School Survey of New York State*, Vol. I, p. 119 (Ithaca, N. Y., 1922).

² Joint Committee on Rural Schools, *Rural School Survey of New York State*, Vol. II, pp. 439 ff and pp. 528 ff.

NEW YORK STATE REGENTS' EXAMINATIONS

PER CENT OF PAPERS WRITTEN ACCEPTED

	1916	1917	1918	1919	1920	1931	1932	1933	1934	1935
English	80.1	74.3	78.3	80.6	81.6	89.7	91.0	89.0	88.5	88.4
Modern languages	66.9	60.9	69.8	73.8	61.0	86.0	86.2	86.3	86.3	89.8
Latin	71.4	70.6	75.1	77.9	73.6	75.4	75.6	81.3	81.4	82.2
Mathematics	65.8	67.4	68.7	63.5	60.5	72.5	69.8	71.3	73.0	72.7
Science	71.0	72.5	69.1	65.3	60.8	83.4	82.6	86.2	85.4	79.9
History ¹	77.5	81.7	86.1	82.2	85.6	85.6	87.9	87.4	85.8	88.8
Commercial subjects	61.9	65.9	73.5	67.4	66.6	80.7	76.2	77.2	76.4	79.6
Drawing ²	75.5	75.1	80.3	82.6	83.0	84.6	78.8	71.2	79.1	83.8
Music ²	75.9	76.7	78.2	85.0	71.6	80.3	81.1	80.7	73.2	78.4
Average	72.2	70.6	73.9	73.7	70.7	81.3	81.1	82.7	82.4	83.5

PER CENT OF PAPERS CLAIMED ACCEPTED

	1916	1917	1918	1919	1920	1931	1932	1933	1934	1935
English	91.7	87.4	90.2	92.7	93.2	97.9	98.4	97.7	98.1	97.4
Modern languages	83.2	79.5	88.9	89.9	79.8	97.8	98.2	98.7	97.9	98.6
Latin	92.0	88.3	91.3	92.8	92.0	95.4	96.5	96.8	96.6	96.1
Mathematics	93.2	93.2	91.4	93.3	89.6	96.8	96.4	97.5	96.1	95.7
Science	87.8	85.7	83.7	80.0	77.7	98.4	98.4	98.9	98.4	97.9
History	90.0	90.5	93.2	94.6	93.8	97.9	97.9	97.5	97.1	98.2
Commercial subjects	83.3	84.9	88.8	85.7	84.7	96.6	94.1	95.0	95.2	96.4
Drawing	88.5	86.0	92.0	94.3	91.8	98.4	88.0	82.5	86.9	92.6
Music	94.9	94.9	90.4	94.3	91.2	91.6	90.3	90.0	86.0	91.8
Average	89.3	87.9	90.4	90.5	88.8	97.4	97.3	97.5	97.0	97.4

responsible for the stability of the results, the data at present available do not afford an adequate basis for judgment any more than the percentage of passing marks is a relevant guide to the reliability or validity of the results. The question may well be asked, however, whether a system which involves the marking of more than a million papers can be justified when the ratings of teachers and examiners are only less than three per cent from attaining perfect agreement. If the approximation in ratings is due to the increase of objective tests in the examination, there may well be found here a sound argument for the abandonment of a system which yields so little. The time and energy now devoted to the mechanics and administration of the examination system may well be used for a more extensive scheme of constructive inspection, one of the functions of which would be to see that the educational values which undoubtedly inhere in the essay type of examination are not neglected.

THE COLLEGE ENTRANCE EXAMINATION BOARD

The College Entrance Examination Board, unlike the Board of Regents of New York State, was established in 1900 as a voluntary organization in

¹ History was included under Social Studies in the later period.

² Drawing later was listed as Comprehensive Art, and Music as Comprehensive Music.

which representatives of secondary schools and colleges could cooperate to harmonize the relations between secondary schools and colleges, to conduct examinations, and to award certificates on the basis of which colleges would admit applicants. It was not intended by the institution of these examinations to prescribe or limit the work of the secondary schools or the requirements for admission to colleges so much as to establish certain recognized standards in both these fields. President Butler, who with President Eliot had been most influential in the creation of the Board, thus defined the Board's functions after fourteen years of its history:

It goes without saying that the college admission examination is not an end in itself, but only a means. It is one of several means—but, I think, an indispensable one—of ascertaining whether a given student is fitted to enter a college and to profit by college instruction and college training. If the Board is able, as I confidently believe it will be, to proceed for another decade on its present lines, it will have brought about everywhere a new spirit of comity and cooperation between secondary school and college teachers, and a new and very helpful understanding of the problems which attend the training of boys and girls who are about to be graduated from secondary schools.¹

In the formal announcement of the organization of the Board it was stated that “the manifest advantages of the examinations held by the Board” were uniformity in subject matter and administration; convenience to the candidates since they were to be conducted at one and the same time at many local centers; cooperation between schools and colleges without sacrifice of individuality; definition of uniform standards for secondary school work; and economy of time, money, and effort in administering college admission requirements.² The Board was authorized by its Plan of Organization and Constitution

to adopt and publish a statement of the ground which should be covered and of the aims which should be sought by secondary school teaching in each of the following subjects (and in such others as may be desirable), and a plan of examination suitable as a test for admission to college: Botany, chemistry, English, French, Greek, history, Latin, mathematics, physics, zoölogy.³

During the first few years following the introduction of the examinations the

¹ College Entrance Examination Board, *The Work of the College Entrance Examination Board, 1901-1925*, p. 131 (New York, 1926).

² *Ibid.*, p. 72.

³ *Ibid.*, p. 73.

Board did not formulate its own requirements but adopted those of existing authoritative organizations, such as the Conference on Uniform Entrance Requirements in English, the American Historical Association in history, the American Philological Association in Latin and Greek, the Modern Language Association in modern languages, and Committees of the National Education Association in mathematics and sciences. On the basis of these requirements "examination questions or other appropriate tests" were prepared in each subject by committees of three examiners (a college teacher as chief examiner, assisted by one college and one secondary school teacher as associate examiners) and submitted for approval or revision to a Committee of Revision consisting of the chief examiners and the five representatives of secondary schools on the Board. A passing mark of 60 on a scale of 100 was set with the provision that no paper was to be finally marked below 60 without a second reading. The actual marking of papers was simple as long as the number of candidates was relatively small (973 at the first examination in 1901); the several groups of readers studied the questions in their subjects and a common standard of marking was reached after a specimen paper was read aloud.

It was soon recognized that to rely on the requirements defined by the national associations was unsatisfactory, since the committees which formulated them were in the main not familiar with the standards to be expected at the secondary school level and were disbanded after they had defined the requirements. In order to provide for continuous study of the working of the requirements and for introducing such changes and modifications as might prove desirable a Committee of Review was created in 1907 not itself to formulate the requirements but to supervise their formulation and revision by specially appointed subject committees of college and secondary school teachers. Such requirements as might be recommended by the committees were not to be adopted by the Board until they had been submitted to schools and colleges for consideration and criticism and all objections had in the opinion of the Committee of Review been eliminated. Every effort was thus made to secure the cooperation of all concerned in the examination to establish acceptable standards in the formulation of the requirements. It should be noted further that the Board began to publish examination questions early in its history and supplemented these later with the publication of *Suggestions and Aids for College Candidates* in various subjects and *Specimen Answers of College Candidates* in English, history, and plane geometry, reproducing answers with marks assigned to them.

More difficult than this task, however, was that of setting uniform stand-

ards of marking in the different subjects. To this end a conference of the chief readers in each of the subjects was held during the week preceding the examinations in June, 1908, "with a view to securing the greatest uniformity of method and point of view in the rating of the answer books." The results in the examinations of 1909, when 62 per cent of the candidates in Latin were marked below 60 and more than half of the candidates in other subjects received less than this passing mark, led in 1910 to the establishment of a Committee on Examination Ratings to unify and systematize the reading and to act as a clearing house for complaints. The function of the Committee was to receive detailed reports from the chief readers describing the methods employed, the results obtained, and the comments of the individual readers. On the basis of information so submitted the Committee was to make suggestions for increasing the accuracy and efficiency both in reading and marking papers and for securing the coopération of the readers in cases of difficulty. To this end the Committee was authorized to conduct conferences of the chief readers and of the readers in the different subjects, but the responsibility was placed on the chief readers to see that the marking was not performed mechanically but in a spirit of fairness and justice to the candidates. Finally, it became the duty of the Committee to study the records of each examination, and, if the results were not satisfactory, as for a long time in the case of history, or if they showed marked fluctuations from year to year, as in mathematics, to discover the reasons.

The Committee on Examination Ratings reported to the Board in 1911 its opinion that unsatisfactory results were due to distinctly difficult questions and to the fact that the examinations were taken in many cases by candidates who were utterly unprepared or very badly taught, or who were unable to secure certificates from their own schools and took the examinations as a forlorn hope. The Committee recommended that the chief readers be given greater authority and responsibility, that through conferences, before the reading of papers is begun, they agree on general principles of procedure to bring about greater unity of work, and that the readers be instructed "to use discretion and to mark the books according to a standard of fairness and common sense, rather than on a strictly mechanical basis" when they found the examinations too long for the time allowed or the questions unreasonable.

The precautions taken to eliminate carelessness, unfairness, and mistakes may be illustrated by the report of the chief reader in history for 1914. Questions were prepared by each member of the committee of examiners in the fall, and were submitted for criticism to the other examiners; the criticisms

were returned and used as the basis for the preparation of a second paper. On the basis of the second papers a trial paper was constructed by the chief examiner, submitted for criticism, and a final paper prepared at a conference of all the examiners. As far as possible no question was admitted which was not dealt with in the customary textbooks. The questions for 1914 were also passed upon by the Committee of Revision, consisting of the chief examiners in all subjects and secondary school representatives appointed by a number of standardizing and accrediting agencies. For purposes of marking the examination papers the committee of fifteen readers was divided into four small groups, each reader marking the same five or six papers in order to maintain a common standard; every paper marked between 65 and 35 was read twice by one college and one secondary school teacher and a final mark assigned after discussion between the two. The function of the chief reader was to be present at the discussion by each group before a standard was agreed upon, to watch for too high or too low marks, to reread debatable papers, to settle doubtful points, and to supervise the marking in general.

The same meticulousness was developed and became the established practice in marking examination papers in all subjects,—study of the questions, determination of numerical values, discussion of types of possible answers, reading, marking, and commenting upon a number of specimen papers, reading some of the most typical papers aloud and discussion by the whole group, pairing readers in doubtful cases, and submission of cases of disagreement to the chief reader. But even these methods did not always succeed in eliminating the difficulties of reliable marking, particularly in the case of inexperienced readers, as was illustrated in the report of the chief reader in geometry in 1922.

There was one striking case to illustrate the danger of over-confidence in any routine system of grading. A book to which one of the inexperienced readers had given a rating of 33 was referred to the committee on hopeless failures. They succeeded in discovering a considerable amount of sound sense behind a smoke-screen of confusing mistakes, and the grade was finally raised to 67. It was an extreme case on both sides. The original mark was too low, because the first reader, proceeding with routine speed, had quite pardonably overlooked a certain amount of meaning which was fairly effectually concealed. The final mark was undoubtedly too high, because it gave the candidate the benefit of the doubt on a number of points, in some instances surely beyond his deserts. On the knowledge of geometry shown, the candidate probably deserved to pass, and it did not seem to be worth while to spend more time in debating his numerical score. A few

other instances, perhaps a dozen among the thousands of books read, showed that re-reading does eliminate some serious errors.¹

In the light of twenty-four years of experience with the examinations conducted by the Board, its secretary, Professor Thomas S. Fiske, analyzed their purposes as follows:

The Board undertakes to ascertain in regard to the boys and girls wishing to enter college not merely whether they have stored in their memories the contents of the textbooks in general use but primarily whether they have made a start, have actually gone a little distance on the road from childhood to intellectual maturity.

Up to the present time the examinations of the College Entrance Examination Board have been, above all, a test of the following qualities.—(1) Power of expression; (2) Intelligent appreciation; (3) Ability to reassemble information; (4) Courage to form and express independent judgments; (5) Concentration, or power to sustain a mental effort. The Board has at all times accepted the principle that it was its province to test the individual candidates in respect to certain clearly defined abilities or acquirements, leaving it to the several universities, colleges, and scientific schools immediately concerned to administer such supplementary tests as should seem to them necessary or desirable.²

The examinations of the Board cannot, in other words, be accepted as the sole evidence of fitness to profit by college education, but the colleges can and do supplement the standing of candidates in these examinations with other records obtained from the schools which they attended, such as reports on ethical behavior, physical health, powers of observation, mental

COLLEGE ENTRANCE EXAMINATION BOARD

RESULTS OF EXAMINATIONS DURING DECADE 1910-1919³

<i>Subject</i>	<i>Number of Answer Books</i>	<i>Percentage of Books Rated 60-100</i>
Greek	8,048	66.1
French	31,602	61.9
Latin	71,496	58.1
Physics	11,079	54.9
Chemistry	6,441	52.1
Mathematics	78,232	51.1
German	23,207	49.7
English	44,136	45.2
History	28,536	35.9
<i>All Subjects</i>	307,865	52.3

¹ *Ibid.*, p. 192.

² *Ibid.*, p. 209.

³ *Ibid.*, p. 167

alertness, ability to participate successfully in coöperative efforts, or team work, skill in laboratory work, and facility in conversation in foreign languages.¹

In 1920 a disturbing factor began to be recognized and to be given serious attention—fluctuations in the results of examinations. The results of examinations during the decade 1910–1919 showed obvious fluctuations in results in different subjects which were as shown in the table on page 48.

The percentage of examination books rated 60–100 in the different subjects for the five-year period from 1916 to 1921 was as shown in the following table:

COLLEGE ENTRANCE EXAMINATION BOARD						
PERCENTAGE OF BOOKS RATED 60-100 AT ORDINARY EXAMINATIONS ²						
Subject		1916	1917	1918	1919	1920
<i>English</i>						
A.	Reading					51.7
B.	Study					39.4
AB.	Reading and Study					46.8
1.	Grammar and Composition	40.7	42.1	54.8	57.1	
2.	Literature	43.2	35.4	41.6	50.6	57.3
<i>History</i>						
A.	Ancient	36.7	37.3	35.2	33.9	36.2
B.	Medieval and Modern	22.7	32.2	35.6	25.1	43.5
C.	Modern			50.0	44.3	44.6
D.	English	23.5	42.4	46.5	35.6	44.4
E.	American					32.7
F.	Civil Government	16.9	47.7	56.5	46.0	44.1
G.	American History and Civil Government	21.0	50.4	38.6	34.4	49.2
<i>Latin</i>						
1.	Grammar	58.1	57.6	43.9	43.8	40.6
2.	Elementary Composition	64.8	58.2	57.5	69.4	65.5
3.	Second Year	68.5	50.6	71.1	60.6	69.3
4.	Cicero and Sight Reading	57.3	53.6	61.8	58.1	50.7
124.	Latin 1, 2, and 4 combined	...	54.9	55.4	53.6	43.3
5.	Virgil and Sight Reading	75.2	68.2	59.8	62.2	57.6
6.	Advanced Composition	56.3	55.1	58.5	64.0	57.2
P.	Sight Translation of Prose	12.5	40.9	49.0	62.6	54.8
Q.	Sight Translation of Poetry	33.3	82.5	42.9	75.6	71.2
<i>Greek</i>						
A1.	Grammar	38.0	58.3	51.8	69.8	54.1
A2.	Elementary Prose Composition	47.8	71.7	57.8	72.3	66.3
B.	Xenophon (Anabasis I-IV)	60.3	66.7	46.4	38.9	50.0
C.	Homer (Iliad I-III)	58.7	68.3	76.6	84.2	96.4
F.	Prose Composition	48.8	56.2	71.4	80.0	57.1
G.	Sight Translation of Prose	57.5	100.0	80.0	100.0	33.3
BG.	Xenophon and Sight Prose	76.1	83.7	85.0	86.6	69.1
CH.	Homer and Sight Poetry	74.6	77.2	82.9	93.2	96.7

¹ *Ibid.*, p. 191.

² College Entrance Examination Board, *Twentieth Annual Report*, pp. 25 f. (New York, 1920).

<i>Subject</i>		<i>1916</i>	<i>1917</i>	<i>1918</i>	<i>1919</i>	<i>1920</i>
<i>French</i>						
A.	Elementary	75.0	59.7	59.5	68.2	67.9
B.	Intermediate	73.2	42.3	51.6	60.7	58.2
BC	Intermediate and Advanced	44.3	56.1	53.1	53.5	61.1
<i>German</i>						
A.	Elementary	55.3	49.8	50.9	19.2	67.4
B.	Intermediate	60.6	35.9	37.5	37.9	70.5
BC.	Intermediate and Advanced	47.1	55.1	43.5	52.6	71.4
<i>Spanish</i>		22.8	39.2	47.7	50.2	57.7
<i>Mathematics</i>						
A.	Elementary Algebra	38.2	63.3	74.7	38.7	73.9
A1.	Algebra to Quadratics	51.5	57.8	46.0	34.6	52.8
A2.	Quadratics and Beyond	46.1	57.8	80.5	46.6	80.8
B.	Advanced Algebra	38.2	74.3	65.5	50.8	62.9
C.	Plane Geometry	38.0	60.6	55.6	35.5	61.5
D.	Solid Geometry	26.1	51.2	58.3	53.0	62.2
CD.	Plane and Solid Geometry		65.5	49.6	31.9	60.9
E.	Trigonometry	60.8	76.7	61.1	69.9	87.5
F.	Plane Trigonometry	55.8	53.2	56.0	56.4	68.6
<i>Biology</i>		48.3	55.5	49.5	52.6	63.3
<i>Botany</i>		70.8	69.4	58.6	41.2	35.5
<i>Chemistry</i>		42.3	50.9	52.2	49.5	59.7
<i>Drawing</i>						
	<i>Freehand</i>	43.0	26.4	25.9	33.9	25.5
	<i>Mechanical</i>	27.5	36.4	41.8	34.5	31.1
<i>Geography</i>		50.0	34.8	18.4	34.8	27.0
<i>Physics</i>		49.8	59.6	52.9	59.0	56.2
<i>Zoology</i>		62.5	66.7	50.0	47.3	26.3

The criticism of the fluctuations in the results of the examinations, once started, continued to disturb the Board for many years. Efforts were made to explain and to justify them. It was stated that the Board from the start had never intended to have a fixed passing mark but to leave to the colleges the decision on the acceptable minimum rating in each subject after the examinations had been held. Other suggested explanations were that with the constant increase in the number of candidates presenting themselves for the examinations there were many who under the new educational conditions were not accustomed to written examinations or were taking them for practice, or that fluctuations were due to variations in the quality of the candidates. After a time these explanations were discarded, and the "violent and regrettable fluctuations" were attributed to

the unfortunate wording of one or more questions. The questions have occasionally been too long and complicated. Sometimes terms unfamiliar to the candidate have been employed. Sometimes the meaning of the troublesome question would have been obscure to anyone. Sometimes the question has been capable of two or more interpretations, one of which was not even suspected by the examiners.

On the language papers occasionally a passage has been set for translation which has proved far too difficult for the candidates either by reason of unusual constructions or by reason of an unfamiliar vocabulary.¹

The fluctuations, it was claimed, were at any rate not due to variations in the standards of the readers who have long experience as teachers and readers with well-established standards based upon agreement reached by the different subject groups before the actual reading of papers is undertaken.²

The charge that the examinations lacked reliability was met by raising the question whether a judgment of value in regard to reliability can be founded on the basis of a single examination, since so much depends upon the equality of questions in character and difficulty and the nature of the special topics selected and the mental characteristics (e.g. memory, imagination, alertness, power of concentration, etc.) stressed by the examiners. Examination results are determined by a large number of factors, such as the following: The candidate's effort or exertion to master the subject matter as presented by the teacher; the adequacy of the textbooks used and the competence of the teacher; the candidate's memory, and logical power or ability to reason; his constructive or creative power; his power of expression and concentration; his intellectual alertness; and his facility in writing rapidly and in arranging his work neatly, depending upon experience acquired at previous examinations.³

With reference to the correlation of examination results communications received from the schools concerned testified that the best pupils received the highest rating and the worst the lowest rating.

If a candidate has attended a good secondary school, there is usually a high correlation between his school record and the results of his examinations with the College Entrance Examination Board. Practically all the secondary schools that make extensive use of the Board's examinations testify to the fact that this correlation is high.⁴

¹ *The Work of the College Entrance Examination Board, 1901-1925*, p. 214.

² Professor L. T. Hopkins in a study on *The Marking System of the College Entrance Examination Board*, p. 14 (Cambridge, Mass., 1921), concluded that the assignment of marks from 1902 to 1920 rarely approximated the normal, even when only those pupils recommended by their schools as fully prepared took the examinations, nor were the results due to the influx of unprepared candidates. The irregularities were, he concluded, a very natural result of the method of reading and scoring the papers, the lack of standardization of values and corrections in conformity with the curve of the error. As the best basis for solving the difficulties Professor Hopkins suggested some approximation to the normal curve, since a certain uniformity in the different subjects should be expected.

³ *Ibid.*, pp. 179 f.

⁴ *Thirty-third Annual Report of the Secretary of the College Entrance Examination Board*, p. 4 (New York, 1933). See also *Thirty-second Annual Report*, where the fear is expressed that colleges may follow the logic of this correlation and admit "without examination students who at the end of the high school course stand in the upper tenth, the upper fifth, or the upper fourth of the class" (pp. 2 and 3).

Since it was the opinion of the Board that fluctuations were due to the character and form of the questions set in the examinations, measures were proposed in 1924 to lessen the fluctuations by the following suggested methods:

(1) Material increase in the number of questions asked at an examination. (2) Better distribution of the questions over the whole field covered by the requirement. (3) Exclusive use of questions previously tried out by experiment in secondary schools. Another form of the last suggestion is the proposal that each group of examiners draw its examination questions from a reservoir consisting of several thousand questions all of which have been tested by experiments in the secondary schools.

In a number of subjects undoubtedly the problem of lessening the fluctuations in the results of the examinations would be solved most simply and most effectively by adopting more detailed and more precise definitions of the requirements and by setting examinations strictly conforming thereto.¹

Ten years later a Subcommittee on Questions of Policy which had been appointed in April, 1933, recommended

that the existing definitions of the examination requirements be liberally interpreted as indicating in a general way the nature and extent of preparation considered necessary and not as prescribing any definite form of instruction, method of preparation, or teaching technique.

It was suggested that examiners prepare their questions in such a way as "to describe the individual candidate with the smallest possible error of measurement," the examining policies to be supervised, reviewed, and coordinated by a reconstituted Committee of Revision. This recommendation was based on the recognition by the Board that there is a problem of examining which needs to be solved by an analysis of the examinations subject by subject "in order to discover methods of increasing their validity and reliability." Further, it was suggested that the examinations should not be used primarily to distinguish between those who pass and those who fail but as usable instruments for placement after entrance into college. The problem of fluctuations which still continued² could be overcome by the Subcommittee's recommen-

¹ *The Work of the College Entrance Examination Board, 1901-1925*, p. 215. The last proposal would, of course, have been contrary to the original intention that the Board's examinations should refrain from assuming any kind of institutional control over the work of the secondary schools.

² In 1931 only 57.6 per cent of the candidates obtained marks above 60 in one of the French examinations as compared with 81.9 per cent in the following year; in 1927 81.9 per cent received above 60 as compared with 52.8 per cent in the previous year. "The complete tabulation of all variations in the percentage of answer books rated between 60 and 100 over a period of years shows striking variations in this proportion from year to year. In the face of such eccentric results it seems more conservative to assume that the populations are comparatively stable and that the examinations vary in difficulty than to assume that the examinations are constant and that the selection of pupils and the quality of the

dation "to re-scale marks in all subjects so that these marks will be more true to form year in and year out." The intention of this recommendation was that candidates should be arranged in the best possible order of merit with their reported marks. It is to be noted that the practice of grouping candidates in percentiles had been tried in 1922 but was pronounced futile in the following year because of the large number of papers marked 60.

The new Committee on Revision was appointed in October, 1934, to "consider all phases of examinations including their relative importance and use, the advisability of continuing certain examinations, the type of examinations to be held, and their number." It is difficult to anticipate the future trends which the Board will follow in the conduct of its examinations. That it has performed an important function and made a significant and much-needed contribution to the institution of acceptable standards in secondary education is clear. The last fifteen years of the history of the Board has been a period of external criticisms and internal self-investigation. This period has coincided with the more extensive development of tests and measurements, the application of statistical methods to problems of education, and an unparalleled increase in the number of students in high schools and colleges. New theories and philosophies of education, the result in part of changes in interpretation and psychology, and in part of rationalizations on the basis of the intellectual maladjustment of large numbers of students, have led to demands for greater flexibility in the range of studies provided and broader adaptations to individual differences of abilities and interests. Many colleges, which in the past jealously guarded the portals of admission, have shown a willingness to accept students on the basis of their secondary school records and the recommendations of their teachers.¹ Finally, experiments have proved that aptitude tests, properly constructed, have greater reliability and higher

preparation vary." The Recommendations of the Subcommittee on Questions of Policy in the *Thirty-fourth Annual Report of the Secretary of the College Entrance Examination Board*, p. 12 (New York, 1934).

¹ Harvard, Yale, and Princeton conducted their own entrance examinations until 1915, when their representatives requested the College Entrance Examination Board to prepare and conduct what came to be known as the New Plan and later Plan B (in contrast to the Old Plan or Plan A) examinations. Under Plan A candidates are required to take enough examinations to make up a complete set of admission requirements aggregating fifteen points, under Plan B candidates who have an acceptable school record are excused from taking examinations in more than four subjects. In 1919 a number of women's colleges—Bryn Mawr, Mount Holyoke, Smith, Vassar, and Wellesley—adopted the same plan.

Many of these institutions have within the last few years indicated their readiness to dispense entirely with entrance examinations and to accept a percentage of candidates on the recommendations of their schools and their records. On the predictive value of school records as compared with examination results see A. D. Whitman, *Value of Examinations of the College Entrance Examination Board as Predictions of Success in College* (New York, 1926).

predictive values than a series of examinations which take three or four times as long to write. To these facts should be added the opposition of progressive educators to external examinations of any kind and the demand of a few secondary schools to be permitted to reorganize their curricula and courses of study without reference to examination requirements on condition that colleges would agree to admit their students on the recommendation of the schools.¹

As a consequence of the rapid spread and use of psychological tests immediately after the War,² the Board in 1922 and again in 1924 adopted the following resolutions:

Resolved: That the College Entrance Examination Board regards with favorable interest the use of general intelligence examinations as supplements to other examinations and stands ready to cooperate in giving general intelligence examinations as soon as giving such examinations may be desired by college members of the Board and as soon as the same may be shown to be feasible.

Resolved: That the Chairman of the Board be authorized to appoint a commission to consider the question of the desire for intelligence examinations and the feasibility of the proposal that the Board participate in their conduct.³

The Commission, appointed in 1924, reported that, on the basis of the fact that many colleges were already using psychological tests either at entrance or afterward and following the opinions of leading psychologists, expressed itself as ready to recommend to the Board the establishment of psychological tests. A committee of experts was appointed in 1925 to prepare examination questions and instructions to supervisors, to score the answers, and to interpret the results. Accordingly in 1926 the Board for the first time adopted the use of "scholastic aptitude tests" but with the reservation that such tests should be supplementary only to other records of candidates for entrance to college. As explained by the committee of experts, appointed in 1925,

Boards of Admission to colleges, now forced to estimate the future worth of candidates, need all the information which is available and pertinent to reach wise decisions. This additional test now made available through the instrumentality of the College Entrance Examination Board may help to resolve a few perplexing problems, but it should be regarded merely as a

¹ W. M. Aikin, "The Purposes of the Eight-Year Experimental Study." *Educational Record*, January, 1935, pp. 107 ff

² See pp. 75 ff, below.

³ *The Work of the College Entrance Examination Board, 1901-1925*, p. 219.

supplementary record. To place too great emphasis on test scores is as dangerous as the failure properly to evaluate any score or mark in conjunction with other measures and estimates which it supplements.¹

The chairman of the committee of experts, Professor Carl C. Brigham, who in 1930 became Associate Secretary of the Board, in a subsequent report offered a further explanation of the policy of the Board in recommending that the scholastic aptitude test be supplementary to other records. He pointed out that in the desire to eliminate some of the elements of unreliability by objectivity of scoring and to rate the individual more fairly "the objective testers have plunged in with the individual rather than the institutional point of view." In other words, they have neglected questions of the content of the curriculum, or the effect of examinations on the curriculum. Further, Professor Brigham argued that as recently as 1933 there were not available tests in controlled language composition, while in a subject such as science, tests "run counter to all legitimate purposes of science study." Professor Brigham suggested that some of the virtues of objective tests could be incorporated in the old type examinations. Of these virtues he selected two, scaling of questions in difficulty in such a way that every candidate can find his own level in a given amount of time, and heterogeneity in the use of larger numbers of questions so that subject matter might be sampled more widely and broader areas of information and reasoning be brought into play. The conflict to be settled was whether requirements for examinations should be defined as a form of institutional control or whether examinations should be used as devices for rating individuals. The problems confronting the Board must be studied experimentally in the same way as the Commission on Scholastic Aptitude Tests had attacked every phase of its problems experimentally. A new need has to be met which had not existed when the Board was first created and that is the provision of as much reliable information as could be secured about a candidate to be used not only for admission to college but for his placement after admission.

An organization set up for the sole purpose of collecting tickets at the gate is now asked to show people to their seats. The notion of a general admissions ticket is yielding to the notion of a more exact description of the individual which will make possible his proper placement in definable universes of knowledge.²

¹ *Ibid*, pp. 44 f.

² *Thirty-third Annual Report of the Secretary of the College Entrance Examination Board*, pp. 13 ff. The whole section is based on a report of the Associate Secretary to the Executive Committee of the Board, pp. 8 ff.

At present the results of these considerations have been to secure a compromise between the new and the old types of examinations, to encourage efforts to make the scoring as reliable and valid as possible, and to use the Scholastic Aptitude Tests to supplement the regular examination. Following the Recommendations of the Subcommittee on Questions of Policy two commissions were appointed to consider the problem of examining, one in history and the other in science. The first of these two commissions was appointed

to study the possibility of a comprehensive examination in history based upon a continuous four-year course and to consider the advisability of setting in each branch of history an examination consisting partly of objective questions and partly of essay-type questions ¹

The second commission was requested to consider the possibility of a comprehensive examination upon a continuous four-year course in science and as alternative to the usual examination in physics a three-hour examination "devised to test the extent to which the student has brought to the enrichment of a study of physics the facts and principles learned in earlier studies of biology, chemistry, or astronomy." A Commission on Examinations in Mathematics had already been appointed in 1933 and in 1935 made recommendations looking to the construction of a new series of examinations to test mathematical aptitude as well as mathematical knowledge.

These developments indicate a definite recognition of the existence of weaknesses in the traditional system of examinations. The experiments upon which the Board has embarked—the retention and improvement of the traditional examination, the tentative introduction of objective tests, and the use of the Scholastic Aptitude Test—will in time provide a body of facts and information on the basis of which the Board may some day be in a position to abandon a caution, which may be laudable at the launching of new methods, in favor of a plan which has stood a variety of tests.

The problem of examinations as far as admission to the college from the secondary school is concerned has been clearly analyzed in the *Thirty-third Annual Report of the Secretary of the College Entrance Examination Board* (pp. 3 ff.):

When a student arrives at the end of his course in the secondary school and applies for admission to college, those who are interested in his career are likely to ask themselves the following questions:

1. How much has he learned?
2. How much more can he learn?
3. How much more will he learn?

The first of these questions may be answered by an examination, although the result may not throw light on a candidate's "originality, imagination, or ability to meet a new situation." The Scholastic Aptitude Test has for its purpose the discovery of a candidate's power to learn, his power of concentration, ability to comprehend or interpret, and other mental gifts. The third question, How much more will he learn?, cannot be answered reliably, because the answer depends on an individual's character, habits, and inclinations.

In the opinion of many teachers it is fairly easy to find out how much a candidate has learned, somewhat difficult to find out how much more he can learn, and almost impossible to prophesy how much more he will learn.

Thus after an experience of thirty-five years the conclusions of the Board coincide with the conclusions of those who have attacked the problem of examinations by scientific, statistical methods. As will be shown in a later chapter there is general agreement among those who have employed these methods that no single test or measure has completely reliable predictive value, that any type of test must be supplemented by a carefully controlled system of cumulative records, and that guidance must be made an essential instrument of education, if students are to be expected to work in accordance with the indications of tests as to their abilities and powers of growth.¹

LATER CRITICISMS OF EXAMINATIONS AND THE "SCIENTIFIC MOVEMENT" IN EDUCATION

At the opening of the century discussions on examinations were still on the basis of opinions. Those who favored examinations claimed that they had an educational value of a high order. As a test of power, it was argued, preparation for examinations trained students to deal with new material, to discriminate between the important and the unimportant, to appreciate the relations of hitherto unrelated details, to grasp a subject as a whole and to combine parts into a vital organic unity, to hold knowledge ready "on demand," to think for oneself. Through examinations the teacher obtained a most impartial estimate of what a student knew, and the student discov-

¹ The effect of examinations like those of the Regents and College Entrance Examination Board upon the work of the secondary schools, the extent to which they influence standards, and the degree to which they advance or retard progress in the revision of subject matter content has not been considered here. This aspect of examinations has been discussed by L. H. Whitcraft, *Some Influences of the Requirements of the College Entrance Examination Board on Mathematics in Secondary Schools of the United States* (New York, 1933), and W. W. Knox, *A Study of Some of the Influences of Regents Examinations in Chemistry* (Albany, N. Y., 1933).

ered what he had really mastered. The timid student acquired confidence and the conceited student gained humility.¹

On the other hand, it was admitted that, while examinations are a great aid in forming habits of mind which tend to make students more accurate, thorough, and clear, they may prove a great curse in setting up false standards. Unless examinations are relegated to their proper place, they encourage cramming and discourage the formation of good habits of work. They could be put in their proper place as a method in training to study, if a student were told that at no stage is success in passing to be the sole test for promotion or admission to college, "regardless of the quality of his previous work or of the method of his preparation." College admission should be based on examinations and student records as reported by the teachers; otherwise examinations tend to crowd the curriculum and are barbarous and worrying for the well-prepared student.²

President Arthur T. Hadley of Yale University argued not for the elimination of examinations but for their improvement. Examinations should fulfill two functions, one looking to the past—a stocktaking, as it were, of what the pupil has studied and mastered, and the other looking toward the future—a test of power to go on. There is a danger, however, that examinations may test only memorization and not thinking or power. The problem is to devise an examination which tests both knowledge and power, and at the same time does not deprive the schools of their freedom. It might be solved by making the range of questions wider, by supplementing the examination with other tests, as, for example, the scrutiny of notebooks and other previous work, and by utilizing the teachers' recommendations. The abandonment of the examinations would remove a stimulus and standards for work. President Hadley suggests a combination of methods, including an examination in essential foundations, such as English and mathematics; greater latitude in auxiliary subjects such as English literature and languages; and acceptance of certificates for studies necessary for general education. The tests of power to pursue further education should be in the hands of those who are to have the students, and the tests of attainment should be in the hands of those who have had them. There would still remain the problem of keeping instruction in the non-examined subjects up to standard, and of devising tests of power rather than of knowledge.³

Four years after this article appeared the traditional methods of discussing

¹ Angie C. Chapin, "Examinations, an Apology," *Educational Review*, Vol. 20, 1900, pp. 519 ff.

² E. H. Nichols, "The Influence of Examinations," *Educational Review*, Vol. 19, 1900, pp. 413 ff.

³ A. T. Hadley, "The Use and Control of Examinations," *Educational Review*, Vol. 21, 1901, pp. 286 ff.

examinations by exchange of opinions began to be attacked. In an article on "Examinations, Grades and Credits" Professor J. McKeen Cattell, who with his students at Columbia University had already begun to apply statistical methods to the study of human traits, threw out the following challenge, which laid the foundations for the subsequent study of the problem of examinations and still later for the development of objective methods of testing.

In examinations and grades which attempt to determine individual differences and to select individuals for special purposes, it seems strange that no scientific study of any consequence has been made to determine the validity of our methods, to standardize and improve them. It is quite possible that the assignment of grades to school children and college students as a kind of reward is useless or worse; its value could and should be determined. But when students are excluded from college because they do not secure a certain grade in a written examination, or when candidates for positions in government service are selected as a result of a written examination, we assume a certain responsibility. The least that we can do is to make a scientific study of our methods and results.¹

Professor Cattell, having in mind the recent creation of the College Entrance Examination Board, feared that the systematization of college entrance examinations would be harmful to the schools, since

It seems scarcely possible to determine whether students are fitted for a college course by means of a written examination. . . . To devise and apply the best methods of determining fitness is the business of the psychological expert, who will probably represent at the close of this century as important a profession as medicine, law or church.²

The article then proceeded to discuss the distribution of grades on a normal curve, Professor J. Rendell Harris's dispersion, and the need of standardizing grades.

When Professor Cattell's article appeared Professor Edward L. Thorndike was already at work in studying the reliability of marking in the examinations of the College Entrance Examination Board. While he welcomed the establishment of this Board because it made the practices of many colleges more efficient and instituted an authoritative body to make the relations between colleges and secondary schools more rational, Professor Thorndike, on the basis of a study of the records of students at Columbia College who had been admitted on the results of the Board's examinations in 1901, 1902,

¹ J. McKeen Cattell, "Examinations, Grades and Credits," *Popular Science Monthly*, Vol. 66, 1905, p. 367.

² *Ibid.*, pp. 368 and 369.

and 1903, showed that "even so carefully managed examinations as these are extremely imperfect means of estimating an individual's fitness for college."¹

The investigation, conducted by Professor Thorndike, showed that success in college cannot be estimated with sufficient accuracy from the entrance examinations. Although students may be admitted on the same average mark, the average "in each case is based on a wide distribution of marks." The estimate at the examination was wrong in 47 out of 50 cases as to success in the junior year. There were clear evidences that the entrance examinations were inadequate as measures of fitness for college. Thus a student might fail at one sitting and pass a few months or a year later as a result of coaching and not necessarily of capacity. In 150 repeated examinations the second mark for the same student showed a median difference of over 22 on a scale of 100, a difference greater than that between the marks of different students chosen at random.

The entrance marks often utterly misrepresented the fitness of a student for college work. Ten students out of one hundred and thirty obtained A's in their junior year, although their average marks at entrance were in some cases in the lowest tenth of the one hundred and thirty and barely above the passing standard. If the passing grade had been slightly higher, one of the very best students of the three college classes would have been excluded. Of those who failed a large percentage would have done better than one-third of those admitted. From this Professor Thorndike concluded that

Sooner or later there will be someone so barred out who would, if admitted, have been the best man in his class. It is a moral atrocity to decide the fitness of an individual for college by a system which, when required to work to a moderate degree of accuracy, is wrong forty-seven times out of fifty. . . . [The traditional examinations] do not prevent incompetence from getting into college; do not prevent students of excellent promise from being discouraged, improperly conditioned, or barred out altogether, do not measure fitness for college well enough to earn the respect of students or teachers; and do intolerable injustice to individuals. There is surely room for improvement.²

Professor Thorndike then proceeded to suggest a method which would be a positive force for selecting those who deserve further education, for promoting such cooperation with secondary schools as would improve conditions and quality of work, and for securing a rigorous, just, rational, and direct

¹ E. L. Thorndike, "The Future of the College Entrance Examination Board" *Educational Review*, Vol. 31, 1906, p. 470. Another report on the same investigation appeared in *Science*, Vol. 33, 1906, pp. 839 ff.

² *Ibid.*, pp. 473 and 474.

measure of fitness for college. He recommended the accrediting of secondary schools on the basis of the actual success in college of candidates indorsed by the schools. Professor Thorndike's interest at the time of writing this article was, as it is at the present time, both in the technical aspects of examinations as well as in their social implications.

It is more important, he wrote, to give advanced education to one boy who most needs it, can profit most by it, use it in the world's service than to prevent from entering upon it a hundred boys who are not able to measure up to its demands.¹

EXPERIMENTS IN THE RELIABILITY OF MARKING

Both Cattell and Thorndike expressed doubts on the validity of entrance examinations to select students for college; inevitably they cast suspicion also on the reliability of marking. The reliability of marking or grading in general was made the subject of investigation by Professors D. Starch and E. C. Elliott, who undertook to determine the range and variation, and the reliability of marks assigned by teachers to examination papers written by high school students in English, mathematics, and history. The experiment aimed to eliminate as many causes of variability as possible by having the same paper marked independently by different teachers. English was selected for the first experiment because it was thought that personal differences were more apt to appear in estimating the merits of a long paper. Two answer papers written by two pupils at the end of the first year's work in English in one of the largest high schools in Wisconsin were reproduced on the same kind of paper as the original with the handwriting, errors, changes, neatness, etc., and sent with copies of the questions to 200 high schools of the North Central Association with the request that they be graded by the principal teacher of first-year English according to his usual practices and standards. One hundred and fifty-two returns were made but for various reasons ten had to be discarded. Of the remainder 51 schools used 70 and 91 used 75 as the passing grade. The general results of this investigation showed a wide range of variation of as much as 35 or 40 points. Thus one of the two papers, B, which had been given a rating of five points above the passing mark of 70, was failed by 22 out of the 142 examiners; this paper was marked on the average eight points lower than the other, A, yet 19 out of the 142 examiners marked it higher and 23 marked it fifteen or more points lower. The range of marks for paper B, the poorer of the two, was somewhat wider than the

¹ *Ibid*, p. 475. Compare with this his statement at the close of an article on "The Distribution of Education," *School Review*, Vol. 40, 1932, pp. 335 ff.

range for paper A. Teachers in small high schools tended to mark the papers more leniently than teachers in the larger schools, but the range of variation remained the same. Students without teaching experience tended to mark more generously than the teachers who participated in the experiment.¹

In order to meet the objection that personal and subjective factors would tend to affect the marking of an English paper, another investigation of the same kind was conducted with papers in geometry on the assumption that such factors would be eliminated in marking papers in an exact science. As shown by the separate grades and the comments on each question care and attention were given to the marking, and yet the result was the same as in the earlier investigation—an extremely wide variation in the marks assigned, even more striking than on the English papers. The reason for this was that the succession of steps, the use of theorems and definitions, the neatness of drawings, spelling, form, make-up and appearance of the papers, and especially the relative value of each particular demonstration or definition were taken into account in estimating the value of the papers. There were found wide variations in standards in high schools distributed over a large area, and, finally, the marks for a single question varied about as widely as for a whole paper.²

A third investigation was conducted by Starch and Elliott in still another subject—history, with the same result. The variation in marks and their range and distribution was almost identical in the three investigations in English, mathematics, and history; the Probable Error (P.E.) was almost the same in all the subjects—5.4 in English, 7.5 in mathematics, and 7.7 in history.

The investigators concluded that variability of marks is a function of the examiner and the method of examination and not of the subject, and that such variability throws suspicion on the fairness and accuracy of existing methods of evaluating the quality of work done in the schools. Hence the following questions arose: What are the factors that produce such wide divergence in the evaluation of school work? and What can be done to secure greater uniformity and more objective reality? The answer to the first question was that there exists a difference of standards among different schools and different teachers, that teachers differ on the relative values to be placed upon the various elements of a paper, and that differences are due to sheer

¹ D. Starch and E. C. Elliott, "Reliability of the Grading of High School Work in English," *School Review*, Vol. 20, 1912, pp. 442 ff.

² D. Starch and E. C. Elliott, "Reliability of Grading Work in Mathematics," *School Review*, Vol. 21, 1913, pp. 254 ff.

inability to distinguish between closely allied degrees of merit. The answer to the second question was the proposal that scales with a definite number of steps be adopted and that a standard curve or table be prepared showing the number of times each particular step should be assigned in the long run. It was proposed, in other words, that for the traditional type of examinations there be substituted the use of standard tests and scales for measuring efficiency in all subjects, similar to the scales which had been recently developed in arithmetic, composition, and handwriting.¹

These studies confirmed experimentally other investigations which had already been conducted and which showed the existence of great variations in marking among teachers, schools, and colleges, a variation due to the subjective methods employed and the absence of objective standards. In 1908 Professor Max Meyer published the distribution of marks assigned at the University of Missouri by forty different professors in different subjects and the variations in standards as indicated by the percentage of students receiving marks of A, B, C, or F. Thus in philosophy 55 per cent of 623 students received A, 33 per cent B, 10 per cent C, and 2 per cent F, while in one chemistry course taken by 903 students only 1 per cent were assigned A, 11 per cent B, 60 per cent C, and 28 per cent F.² Professor W. F. Dearborn found similar discrepancies in the grading of students at the University of Wisconsin by forty-five instructors in seven subjects;³ he concluded that

marks, representing as they do the teacher's estimate of mental abilities of various sorts, may themselves be naturally distributed according to the same frequencies as are the abilities which they are designed to represent. In so far as the teacher's judgment is correct and is made of a sufficiently large number of pupils, the frequency of the different marks given should be the same as in a "normal" distribution curve.⁴

At Cornell University it was found by I. E. Finkelstein that the same students in a year course received entirely different grades from the two instructors who taught the course in each semester.⁵ Further, as Starch found in an experiment conducted at the University of Wisconsin, the same teach-

¹ D. Starch and E. C. Elliott, "The Reliability of Grading Work in History," *School Review*, Vol. 21, 1913, pp. 676 ff.

² M. Meyer, "The Grading of Students," *Science*, Vol. 28, 1908, pp. 243 ff.

³ W. F. Dearborn, *School and University Grades*. Bulletin of the University of Wisconsin, 1910, No. 368. See also similar results in W. T. Foster, *Administration of the College Curriculum*, Ch. XIII (Boston, 1911).

⁴ *Ibid.*, p. 10.

⁵ I. E. Finkelstein, *The Marking System in Theory and Practice* (Baltimore, Md., 1913).

ers gave different marks when they regraded their own papers without knowledge of their former marks.¹

Later studies only helped to confirm the results of the earlier investigations. Sandiford reports the result of regrading an essay on the same subject set by the English Department of the University of Toronto. "An essay which had secured a mark of 80 in one year was copied by a student in another year and handed in as his work. The mark given to it the second time was 39."² A physics paper for a matriculation examination was passed around after two weeks of examining and marked by the various associate examiners in Toronto; the range of marks given was from 50 to 70 with 60 as the passing mark. In another investigation reported by Sandiford

A paper in Honour Matriculation History was scored 50 (60 less 10 for misspellings) by the examiners. This paper was untidy, illegible, and badly spelled, yet as a history paper it had some merit. It was copied out in fair handwriting and the mistakes in spelling corrected, but all other mistakes were retained. The rewritten paper was again sent to the examiners and scored by them as a new paper. The mark given to it this second time was 70 (70 less 0 mistakes for misspellings). We can only conclude that the examiners were influenced in giving 10 marks less than it apparently deserved by its general untidiness and illegibility. They had, however, been instructed not to deduct marks for any of these extraneous factors except spelling and for this a maximum of 10 marks was the limit.³

Finally, and without exhausting the list of studies on the reliability of marking,⁴ as an answer to the question *Quis custodiet ipsos custodes?* in examinations Professor Ben D. Wood cites the following incident:

The facts of a subjective scale are well illustrated in the following anecdote concerning the grading of history papers by a group of college professors of history in the summer of 1920. One of the five or six expert readers assigned to a certain group of history papers, after scoring a few, wrote out for his own convenience what he considered a model paper for the given set of questions. By some mischance this model fell into the hands of another reader who graded it in a perfectly *bona fide* fashion. The mark he assigned to it was below passing, and, in accordance with the custom, this model was rated by a number of other expert readers in order to ensure

¹ D. Starch, "Reliability and Distribution of Grades." *Sciences*, Vol. 38, 1913, pp. 630 ff.

² P. Sandiford, *Educational Psychology*, p. 303 (New York, 1928).

³ *Ibid.*, pp. 304 f.

⁴ See for further references F. J. Kelly, *Teachers' Marks, Their Variability and Standardizing* (New York, 1914); and G. M. Ruch, *The Improvement of the Written Examination*, Ch. III (Chicago, 1921).

that it was properly marked. The marks assigned to it by these readers varied from 40 to 90.¹

Four factors have been mentioned by Starch as possible explanations of the discrepancies found in the distribution of marks.

- (1) Differences in the standard of severity or leniency in different schools;
- (2) differences in the standard of severity or leniency of different teachers;
- (3) differences in credit or penalty assigned by different teachers to any given fact or error in a piece of work; and (4) minuteness of the discrimination between successive steps of merit or quality in a given scale of qualities.²

The effect of the sustained attacks on the reliability of examinations, coming as they did at the same time as cumulative evidences of the nature and distribution of individual differences, expansion of enrollments in high schools and later in colleges, and movements to bring about changes in curricula and courses of study, was to bring the traditional type of examination and its marking into disrepute. Other reasons also tended to produce the same results. Examinations were criticized as being injurious to the health of pupils and leading to overstrain and worry. They become objectives in themselves and encourage cramming, memorization and faulty methods of study inconsistent with modern aims in education. The brief time allotted for examinations is responsible for careless use of English and poor handwriting. They are on the whole unnecessary and the time devoted to them could more profitably be used for more study, review, or recreation. And, finally, a good teacher can grade his pupils without examinations.³ On the technical side it has been argued that the traditional examinations yield inaccurate measures of the achievements of pupils, since the marking is affected by personal and subjective factors. The traditional examinations include only a few questions which are not of equal difficulty and whose weighting by teachers is subjective.⁴ Since the number of questions is small, the pupils are subjected to chance selection and preparation. The content of examinations, ambiguous, indefinite, and using catch questions, is not, it is claimed, in agreement with sound educational aims. Finally, the rate of work is neglected so that a true

¹ Ben D. Wood, "Measurement of College Work." *Educational Administration and Supervision*, Vol. VII, 1921, pp. 301 ff.

² D. Starch, *Educational Psychology*, pp. 436 f (New York, 1920).

³ C. W. Odell, *Traditional Examinations and New Type Tests*, pp. 10 ff (New York, 1928)

⁴ R. Comin, in an article on "Teachers' Estimates of the Ability of Pupils" (*School and Society*, Vol. 3, 1916, pp. 67 ff) reported a very wide variation in the ranking of twenty-three problems in order of difficulty by twenty teachers; one problem which was considered the easiest by one teacher was ranked twenty-first in difficulty by another.

measure of the ability of pupils working against time cannot be secured.¹

In a Symposium on The Value of Examinations conducted by the *Journal of Education*² in 1929 the opinion was expressed by a large number of the leading superintendents of education in the country that examinations as a test for passing a course, for promotion, or for the award of certificates and diplomas should be discarded. Examinations—and the majority appeared to have the new type tests in mind—are valuable as teaching devices: to encourage study and review, to enable pupils to acquire a grasp of larger relationships, to test the attainment of pupils, and to help teachers to discover the strength and weaknesses in their pupils and to correct the latter. In a word the chief use of examinations is for purposes of diagnosing pupils' difficulties, improving methods of instruction, and guidance. Promotions should be based on a combination of various methods—examinations, records, and teachers' judgments. In all these discussions the reference was not to external examinations but to those conducted by teachers in their own classrooms.

MARKING OF ESSAY EXAMINATIONS

A few attempts have been made to prove that the traditional essay type of examination can be so improved both in form and methods of marking as to yield more reliable results, but these attempts seem to approximate in general concept the new type tests in so far as they call for more restricted answers to more specific questions. Provided the procedure both in formulating the questions and in marking is perfectly straightforward, it is claimed that essay examinations may be marked with reliabilities of .80 and over, as compared with reliabilities from .30 to .50 which are usual, and .70 and over which are rare, where such care is not taken. Methods should be used to restrict the answers to specific questions. But the suggestions that are offered on procedure do not differ from those already adopted by such examining bodies as the Regents or the College Entrance Examination Boards, that is, agreement on what the questions should be marked for, analysis of ideal answers, and assignment of a certain number of points to each significant part of a question.

Thus in an article on "Reliable Reading of Essay Tests" the authors, J. M. and R. C. Stalnaker, conclude that

If essay tests are to be used in achievement examinations (and because of the prevailing belief that they fulfil a function which objective tests do

¹ W. S. Monroe, *Written Examinations and Their Improvement*, pp. 10 ff. University of Illinois Bulletin, No. 9, 1922.

² *Journal of Education*, Vol. 119 (Boston, 1929).

not, there is every indication that they will be), reliable reading of them must be obtained. Examinations will never be valid until they are reliably read. If care is taken to formulate essay-test questions in such a way that a restricted form of answer is required and if readers will judge papers on the basis of specific, predetermined criteria which trial has shown can be consistently evaluated, there is no reason why reliability of reading cannot be measurably improved. Indeed, in the reading of English examinations at the University of Chicago the reliability has improved from .42 in October, 1931, to an average of .92 in November, 1932.¹

The following examples are given by the authors of the formulation of questions to secure a definite and restricted type of answers which should serve as measures not of ability to recall memorized facts but of a specific and narrow ability. Instead of asking students to "Discuss the causes of the World War," the answers to which it would be difficult to mark reliably, the question should be put in this way: "Discuss the causes of the World War with specific reference to (a) the alliances in 1914 and their effects; (b) the rivalries of colonial expansion; (c) militarism in Germany; (d) . . . etc." Or again answers to a question in this form "Compare the writings of Corneille and Racine" can be marked less reliably than if the form were "Compare the writings of Corneille and Racine as to (a) modernity; (b) use of action; (c) observance of the unities, . . . etc."² Suggestive as these proposals are, they are approximations to the objective type of tests, which carry the process of analysis of content further.

There still remains the argument in favor of the essay type of examination that it provides better evidence of understanding, reasoning, and ability to organize information than do the objective tests which, it is objected, measure only knowledge of facts and information. Since this is the most common criticism of objective tests it will be dealt with later. Here it is only necessary to refer to one experiment on the results of essay and objective tests in the same subject (United States history) given to the same pupils. It was found in general that when an essay test and an objective test are carefully made and when the scoring of the essay examination is kept as objective as possible, the tests have about equal merit in measuring the understanding of the pupils. While the differences in correlation of the two tests are in favor of the essay test as an instrument for measuring understanding, the differences are

¹ J. M. and R. C. Stalnaker, "Reliable Reading of Essay Tests" *School Review*, Vol. 42, 1931, pp. 599 ff.

² *Ibid.*, p. 601.

small and are balanced by the time involved in the construction and marking of the tests.¹

The discussion of objective tests in a later chapter will show that those who have attacked the problem of constructing such tests intelligently and with insight into education have endeavored to meet the challenge that objective tests measure reasoning as well as knowledge of facts. The challenge is salutary, for there is a real danger that, because of the greater reliability and validity of sound objective tests, the educative value of writing essays and their value as a method of training in expression, in assembling materials, and in organizing them logically may be neglected, or, if not neglected, be relegated to classes in English composition alone.

¹ R. B. Weaver and A. E. Traxler, "Essay Examinations and Objective Tests in United States History in the Junior High School," *School Review*, Vol. 39, 1931, pp. 689 ff

CHAPTER III

THE SCIENTIFIC ATTACK ON EXAMINATIONS

SELECTION OR DISTRIBUTION IN EDUCATION

THE critical attack on the old type of examinations could not have produced successful results had there not been developing, side by side with the statistical investigations into the reliability of marking, a movement for the construction of more scientifically accurate, reliable, and valid types of examinations which should at once eliminate the subjective factors and establish objectively measurable standards. Evidence had been accumulated that no amount of care in the marking of papers of the essay type—discussion of standards before the marking was undertaken, the reading of papers by two or more examiners, the distribution of students on a normal curve—could successfully eliminate the fallibility of subjective judgments. There was an additional factor which necessitated a system whereby the abilities of pupils and students could be more accurately judged, a factor which has only recently begun to make itself felt in European and other systems of education. This was the remarkable and unprecedented increase in the number of students flocking first into the high schools and later into the colleges. This increase which became more rapid in the last two decades inevitably raised the problem of discovering types of curricula and courses best suited to the ever-widening range of individual differences. If the high school and college were to be accessible to all who wished or were prepared to request admission, it became the inevitable task of educators to discover the kind of education from which individual students could best profit in accordance with their abilities.

Admission to the public high school has never been selective, and increasingly selection for admission to state-maintained universities is beginning to disappear. While this tendency is inevitably accompanied by the danger of lowering standards¹ or of catering to mediocrity, American educators are not altogether oblivious to this danger, and for two reasons. The first is that as contrasted with the European tradition there does not exist in American education as widespread and unwavering an acceptance of a somewhat rigid definition of culture or of scholarship. The American tradition represents sometimes a conflict between, sometimes a blending of the two traditions

¹ See I. L. Kandel, *The Dilemma of Democracy* (Cambridge, Mass., 1933), and J. L. Tildsley, *The Mounting Waste of the American Secondary School* (Cambridge, Mass., 1936).

that have made American culture and civilization what they are: the Western tradition of a liberal education and the pragmatic tradition of the frontier. The second reason is the acceptance of the democratic principle of equality of educational opportunity which in contrast to earlier notions is gradually beginning to be defined as an opportunity to obtain that education from which the individual is most capable of profiting. Elasticity in the concept of a liberal or general education and the emphasis on individual ability and capacity are at once the result and the cause of the movement to establish more accurate measures of educability.

Hence, while it is true that in its origin the attempt was made to discover more valid and reliable substitutes for the traditional examinations, there has gradually developed the recognition that the purpose of examinations of any kind should not be primarily to separate the sheep from the goats, to distinguish between those who should pass and those who should fail, but rather to discover the abilities of pupils and students and to provide for them the kind of education best suited to their abilities. To this purpose some contribution was made by applying to education some of the principles of efficiency which had been developed in industry and business. In so vast and costly an undertaking as the educational system, maintained and supported by public taxation, the question presented itself in these forms: Is there any justification for maintaining such a system if a certain percentage of students fail? Should it not be the duty of educators to devise such forms of education as will prevent failure by adequate adjustments to the abilities of the pupils? Should not measures be taken to encourage persistence in schools or colleges? How much and what kind of education has an individual derived who leaves with a record of failures? Here again the statement must be repeated that the American system of education unlike the European is neither competitive nor selective, nor has education in the past been associated with the idea of social status.

It is perhaps not an exaggeration to say that some systems of education and many teachers seek to prove the maintenance of high standards by pointing to the large percentage of failures. In American education the present tendency is to regard failure as a criticism of the system. This tendency may be illustrated by a number of quotations:

There are no misfit children. There are misfit schools, misfit texts and studies, misfit dogmas and traditions of pedants and pedantry. There are misfit homes, misfit occupations and diversions. In fact there are all kinds

and traditions of misfit clothing for children, but—in the nature of things there can be no misfit children.¹

As well say a man does not fit his clothes as say the child does not fit the school. The child is the standard to which all other things must be adjusted, and to the extent that this adjustment is imperfect, to that extent the school fails.²

The duty of publicly supported educational institutions, especially those that operate under the compulsory attendance law, is to help all pupils, whether they be geniuses, mediocrities or morons. To fail large masses of young people on the ground that they have not lived up to the minimum essentials of a predetermined curriculum is both a professional blunder and a social injustice. There is no way of adjusting the school except by studying the capacities, achievements, interests and needs of growing individuals.³

A uniform standard of attainment for all secondary school boys and girls cannot be maintained. Standards should be determined for each individual on the basis of his abilities and interests within the limits of social value.⁴

In a system of education where growth and development are the key words, success is not to be measured primarily by the accumulation of credits and units, but by the degree and kind of growth along the lines of native capacity and individual needs. The word "failure" will have no place in a philosophy or scheme of education which is so gauged and measured.

It naturally follows that certificates of graduation or diplomas will be granted to all who have made the growth possible for them over a given period of time. The degree of development would be indicated on the certificate. To such as had grown in a certain direction and to a prescribed extent, the certificate would serve as a passport to institutions of higher learning. To others whose growth was of lesser extent or along other lines the certificate would stand as a symbol of recognition for accomplishment and it would prevent a sense of futility and frustration on the part of students who put forth honest effort.⁵

(1) Once a college has admitted a student it has a moral obligation to do

¹ F. Burk, *In re Every Child, a Minor vs Lockstep Schooling* (San Francisco State Normal School, Monograph C, 1915).

² B. R. Buckingham, *Research for Teachers*, p. 299 (Chicago, 1926).

³ B. D. Wood, *loc. cit.*, p. 13.

⁴ Associated Academic Principals, New York State, *Summary Report of the Committee on Secondary School Problems*, p. 16 (1932).

⁵ California State Department of Education, Division of Secondary Education, Bulletin No C-9, September, 1931, *Guide for Counseling in the Secondary School*, pp. 20 f.

everything within reason to help him succeed. (2) The college is much more than an agency for providing academic instruction. (3) The college, peopled by anxious and perplexed adolescents, has an opportunity, and, what is more, a responsibility to help these youths resolve their insistent and portentous dilemmas.¹

Although it is obvious that they (poor students) do not deserve, or rather that their lot will not be improved (and might be aggravated) by receiving passing grades in physics, it is the opinion of a growing number of educators that such students deserve something more constructive and helpful for their time and money than merely a failing grade. It is equally true that society deserves more for the time and money spent on such pupils than the mere stigmatizing of them as failures.²

The implications of these quotations have been summarized by Professor Henry C. Morrison of the University of Chicago in the statement that "Teachers should spend half their time studying their pupils as individuals, and the rest of their time doing what that study shows to be desirable and necessary."³

Examinations from the American point of view (in theory generally, in practice to some extent) have begun to be regarded as instruments and de-

¹ W. H. Cowley, "The College Guarantees Satisfaction." *Educational Record*, January, 1935, p. 29.

² F. Palmer, Jr. "The College Physics Testing Program." *Educational Record*, January, 1935, p. 91.

³ From the European point of view the question that is suggested by these quotations is how much of a genuine education is acquired by the *non-valeurs* and the large percentage of students who fail in the French *lycées* and *collèges*, or those who fall under the second category in *Geist und Torheit auf Prämienbanken* by W. Hartnacke and E. Wohlfahrt (Dresden, 1934) in the German secondary schools, or the 30 per cent of those who fail in the first secondary school examination in English secondary schools.

The quotations given above may be paralleled by the address of Professor C. Delisle Burns at the Conference on Examinations held at Eastbourne in 1931. The reader is referred to the whole address which will be found in the *Conference on Examinations*, pages 226-234; to the quotation cited earlier (pp. 17 f) the following may be added "It is important that we should discover, not merely how far the old-fashioned examination should go—and I do not think it goes very far for our modern needs—but also you must not make success or failure in an old examination system, designed in an age which has gone, the test of the kind of person you need in the modern world. You need a new kind of examination system, one based rather on the social assumption that you want distributions of functions within each section of society. It is much less important to see who is at the top and who is at the bottom. There is no top and bottom really; society is simply an interlocking of certain structures—almost organic structures—in regard to which you cannot say that one part of the structure is purely directive and the other part is purely instrumental. . . We therefore want new kinds of tests, and if we can discover the kinds of test that will give us the sort of banker that is the best banker and the sort of industrialist that is the best industrialist, and the sort of railwayman or railway manager that is the best railwayman and railway manager, the result will be that we shall save the educational system from such a collapse as that of the Chinese educational system when the situation changed too rapidly" (p. 233).

The statement of M. Léon Brunschwig, quoted on p. 16, carries implications of the same kind as do the suggested creation of a *Commission Permanente de Sélection et d'Orientation* in the Ministry of National Education in France, and the discussions of the *baccalauréat* in the French Committee's report, *La Correction des Épreuves Ecrites dans les Examens*, Part I (Paris, 1936).

vices for guiding, advising, and placing students, and for these purposes the instruments have had to be made as accurate, reliable, and valid as possible. Further study, however, has indicated that no single examination or test of ability of any type can be relied upon as the sole device for the educational guidance of students. Hence the research of the past two decades has been devoted to the perfection of all available measures of ability and attainment, and to the discovery of the kind of information that is most needed about an individual in order to supplement existing measures and to yield the best diagnostic and prognostic results. It is perhaps not out of place to state that in the main, possibly because of the necessities of the situation, possibly because of an expectation that the able student can take care of himself, more attention has been paid to the poorer and less able student.

THE SCIENTIFIC MOVEMENT IN EDUCATION

The demands imposed upon educators by the conditions which from the beginning of the century were peculiarly American¹ and the criticisms of the traditional practice of examinations coincided with, if they did not actually stimulate, the development of what has been called the scientific movement in education, that is, methods analogous to those employed in the physical sciences. Sir Francis Galton in England and James McKeen Cattell in the United States had already begun before the close of the nineteenth century to experiment with the application of statistical methods to the measurement of human traits. Thus in 1890 Professor Cattell indicated the scope of the new methods in the following statement:

Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement. A step in this direction could be made by applying a series of mental tests and measurements to a large number of individuals. The results would be of considerable scientific value in discovering the constancy of mental processes, their interdependence, and their variation under different circumstances. Individuals, besides, would find their tests interesting, and, perhaps, useful in regard to training, mode of life or indication of disease. The scientific and practical value of such tests would be much increased should a uniform system be adopted, so that determinations made at different times and places could be compared and combined.²

Sir Francis Galton in his remarks on Cattell's article made the following

¹ Somewhat similar conditions are imposing the same demands upon other systems of education in all parts of the world.

² J. McKeen Cattell, "Mental Tests and Measurements," *Mind*, Vol. XV, 1890, p. 373.

important addendum, which indicated the direction to be taken subsequently by the scientific movement in education.

One of the most important objects of measurement is hardly if at all alluded to here and should be emphasized. It is to obtain a general knowledge of the capacities of a man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. We thus may learn which of the measures are the most instructive. The sort of estimates I have in view and which I would suggest should be noted (? for private use) is something of this kind—"mobile, eager, energetic; well-shaped; successful at games requiring good eye and hand; sensitive, good at music and drawing." Such estimates would be far from worthless when made after only a few minutes' talk; they ought to be exact when made of students who have been for months and years under observation.¹

In the above-mentioned article Cattell proposed ten tests: dynamometer pressure, rate of movement, sensation areas, pressure causing pain, least noticeable differences in weight, reaction time for sound, time for naming colors, bisection of a 50 cm. line, judgment of ten seconds time, and number of letters remembered on one hearing. The methods of measuring sensory discrimination and rapidity of reaction, introduced and further expanded by Cattell and Farrand, were published in 1896, but while they were an important contribution on the technical side, their influence was still limited to laboratory experimentation. The more direct influence on the practical problems of education came at about the same time from two directions—the application of statistical methods to education by Professor Edward L. Thorndike and the development of intelligence tests, primarily for subnormal children, elaborated by Binet and Simon. In 1904 Thorndike published his *Introduction to the Theory of Mental and Social Measurements*; Binet's scale of intelligence, first published in 1905, was introduced to the United States in 1908 by H. H. Goddard, director of the psychological laboratory at the Training School for Feeble-minded Children at Vineland, New Jersey. It is not necessary here to discuss the further development of intelligence tests, their revision, elaboration, and application to normal pupils.²

¹ Sir Francis Galton, *ibid.*, p. 380.

² For this history see National Society for the Study of Education, *Seventeenth Yearbook*, Part II, The Measurement of Educational Products (Bloomington, Ill., 1918), and *Twenty-first Yearbook*, Part I, The Nature, History, and General Principles of Intelligence Testing, and Part II, The Administrative Use of Intelligence Testing (Bloomington, Ill., 1922); R. Pintner, *Intelligence Testing* (New York, 1923); and A. T. Wylie, "A Brief History of Mental Tests," *Teachers College Record*, Vol. XXIII, January, 1922, pp. 19 ff.

While progress was being made with the development and improvement of techniques of measurement and statistical methods in general and educational psychology, the foundations were beginning to be laid for the application of such methods to school subjects. In 1894 Dr. J. M. Rice undertook the preparation of standardized tests in school subjects, starting first with tests in spelling and extending later to tests in arithmetic and composition. The pioneer efforts of Rice did not carry conviction in the educational world until more exact standards of measurement began to be applied to the problem by Thorndike and his students. The *Arithmetic Tests* of Stone (1908) and the *Handwriting Scale* of Thorndike (1909), because of their employment of more scientific procedures than were at the command of Rice, carried immediate conviction. From these early beginnings the development of standardized tests spread rapidly to other fields of instruction, first at the elementary, then at the secondary, and still later at the college level.

There were thus available two instruments—the measurement of intelligence and the measurement of achievement in education—which were hailed as the final solutions for providing the right education for the right pupil, for meeting all administrative problems in education (its appropriate distribution, and the classification and selection of pupils and students, vocational and educational guidance, and the standardization of the educational product), and for correcting the fallibility of traditional examinations. The hopes which were aroused by tests of intelligence were thus expressed by Professor R. Pintner:

Intelligence testing is of recent growth, and the accomplishments of the last fifteen years have been great. Nevertheless we stand merely at the beginning of an important chapter of human behavior. The future will see new and more accurate tests. It will see tests applied for purposes now unthought of. The types of individuals tested . . . will undoubtedly increase. A future edition of such a book as this may well have such chapter headings as "The Immigrant"; "The Voter"; "The Applicant for a Marriage License"; "The Candidate for Public Office"; "The Civil Servant"; and so forth. For intelligence is one of the aspects of an individual's personality that is of great importance in modern civilized life. And now that we are able to measure it with a fair degree of accuracy, intelligence tests will find a wider and wider field of application.¹

These expectations were not only based on a characteristic tendency to accept mechanical devices in American education, but also in a larger measure sprang from the fact that intelligence tests had proved to be a ready

¹ K. Pintner, *Intelligence Testing*, p. v (New York, 1923).

and on the whole successful instrument for the speedy grading of the large body of men needed when the United States entered the World War.¹ It was soon discovered, however, that too much was being claimed for intelligence tests. They were severely criticized as being too deterministic and fatalistic in their assumptions and capable of being employed as instruments with which to attack the American ideal of equality of educational opportunity, if the results were used to determine arbitrarily the scope and limits of an individual's educability.² These criticisms, which were launched by Professor William C. Bagley against the misuse of the intelligence quotient, threatened in part by those who seemed to exaggerate the significance of intelligence testing and in part by inferences drawn from the results of the army tests, had the very salutary effect of directing attention to a definition of what was meant by intelligence which the tests professed to measure. In a report on "Principles underlying the Construction and Use of Intelligence Tests" Professor Stephen S. Colvin had ample grounds for writing

Indeed, there is danger at present that the movement in the direction of intelligence testing may grow out of all bounds; that it may be misunderstood in theory and erroneously and even harmfully applied in practice.³

The determinists were too ready to accept the intelligence quotient as a fixed and unchanging datum. Current opinion on the question may be indicated by Professor Colvin's discussion of general intelligence.

By the word *general* is commonly understood an innate ability or group of abilities that lie at the basis of the acquired intelligence of an individual. *Intelligence itself is not inborn, only the capacity to become intelligent.* . . . While all competent authorities would agree that the expression "general intelligence" designates inborn capacity to acquire intelligence in the various situations of life, they would disagree as to the further interpretation of this term, in regard to the significance not only of "general" but also of "intelligence." There are some who hold that the word "general" signifies a single inborn capacity to become intelligent in all situations; others that the term "general" means nothing more than that a person is born with a large number of specific capacities, more or less related, which enable him to acquire intelligent behavior in many different activities. The supporters of this first view, notably Spearman, Hart, and Burt, explain innate intelligence as a "general common factor." . . . [Thorndike] holds to a multiplica-

¹ C. S. Yoakum and R. M. Yerkes, *Army Mental Tests* (New York, 1920).

² W. C. Bagley, *Determinism in Education* (Baltimore, 1928); *Education and Emergent Man*, Ch. VII (New York, 1934).

³ National Society for the Study of Education, *Twenty-first Yearbook*, p. 11 (Bloomington, Ill., 1922). Quoted by permission of the Society.

ity of innate abilities that are related in varying degrees. He believes that between desirable single traits in a single individual there is a positive relation. . . . After all, to the practical schoolman it makes very little difference whether general intelligence is a central factor or a bundle of different abilities related positively; *the child cannot be treated as a unit—he must be discovered in his various tendencies and abilities* and if we wish to know him as he really is, we must be able to work out the “psychogram” which Professor Whipple has mentioned. . . . [Intelligence tests] determine an individual’s intelligence largely in terms of what he has learned, thus obtaining a measure of his ability to continue learning. . . . We shall not be far from the truth when we define general intelligence as *a group of innate capacities by virtue of which the individual is capable of learning in a greater or less degree in terms of the amount of these innate capacities with which he is endowed. . . . Intelligence must be acquired. Only the capacity is inborn.*¹

Since subsequent developments in measurement have been based on this clarification of the intelligence test, Professor Colvin’s statement may be supplemented by another definition by Professor Ben D. Wood of Columbia University:

The intelligence test is a highly specialized instrument, designed mainly to measure the one factor of mental alertness. It does not, unfortunately, measure the intellect directly, but something else which under certain conditions is almost if not quite synonymous with intelligence. This “something else” that is measured in lieu of intelligence is very complex, and includes such elements as achievement in English language, information in special fields, ability to think in terms of facts and relations in specific fields of information, expressed in a certain language. The analysis might be carried to much greater detail. But obviously, the intelligence test is a direct measure of certain types of achievement and habit, and measures intellect only when each individual measured has had an equal opportunity in those specific types of achievement, except insofar as inequalities in opportunity are due to inequalities in native intellect.²

In other words, the intelligence test itself in any case measures scholastic attainments; it may reveal capacity but not the specific nature of such capacity. Further investigation which resulted from the recognition of this fact also indicated that the results are not stable and unvarying. Finally, it was realized that the test itself does not measure the total multiplicity of innate abilities and characteristics that make up intellectual capacity and personality. From the point of view of prognosis or prediction of scholastic success

¹ *Ibid.*, pp. 11–17. Quoted by permission of the Society.

² Ben D. Wood, *Measurement in Higher Education*, pp. 53 f. (New York, 1923).

it was admitted that intelligence alone, however carefully measured, is only one of a number of determining factors. Intelligence tests have thus been found to measure what they measure and, while useful as a general guide, cannot be employed as measures either of specific types of ability or of the total complex of abilities, habits, and perseverance which are determined by environment, experience, education, and opportunity and which in turn determine the nature and extent of the educational achievement of each individual.

Nor were standardized tests of achievement exempted from criticism. It was recognized from the start that certain subjects, such as arithmetic, handwriting, spelling, and reading, lend themselves more readily to measurement by standardized tests than do others, such as history, geography, and literature. Particularly is it difficult to devise norms for an educational system as varied and diversified in pupils, environment, and teaching personnel as the United States. While from one point of view the introduction of standardized tests was welcome and salutary because of the existence of so many widely differing standards, it was felt that such tests helped to establish norms for groups rather than standards for the individual pupil. For in the words of Sandiford:

A standard differs from a norm. It represents the score which a pupil is expected to make, while a norm is the median score that a sample group makes on a test. The standard to be set must take in the whole programme of the pupil—at home as well as at school. Usually, however, the norm is considered as the standard to be achieved.¹

Another serious criticism which tended to throw suspicion on the whole movement was the fact that tests were too frequently prepared by experts in measurement without the coöperation of subject matter specialists. Equally open to criticism is the tendency on the part of subject matter specialists to prepare their own tests without a thorough realization of the technical problems involved in such preparation. Such criticisms do not, however, affect the general principles upon which standardized tests are based and are valuable in directing attention to the dangers which may result from their misuse.

The arguments in favor of measurement in education have been most clearly and succinctly stated by Professor Edward L. Thorndike:

The task of education is to make changes in human beings. We teachers and learners will spend our time this year to make ourselves and others different, thinking and feeling and acting in new and better ways. These classrooms, laboratories and libraries are tools to help us change human nature for the better in respect to knowledge and taste and power. For mastery

¹ P. Sandiford, *Educational Psychology*, p. 310 (New York, 1928).

in this task, we need definite and exact knowledge of what changes are made and what ought to be made. In proportion as it becomes definite and exact, this knowledge of educational products and educational purposes must become quantitative, taking the form of measurements. Education is one form of human engineering and will profit by measurements of human nature and achievement as mechanical and electrical engineering have profited by using the foot-pound, calorie, volt, and ampere. . . .

It may be expected that measurements of achievements and capacity and their quotients will soon be developed for use in high schools, colleges, and professional schools. It surely is unwise to have the measure of college students' achievement in English composition, or trigonometry, or beginning chemistry, or economics, or second-year French depend upon the caprices of a thousand different individual instructors, if by enough ingenuity and care we can devise tests that will measure their achievements uniformly and precisely. The present condition at its best is shocking. The average correlation between grades given in a subject and a student's real achievement in it is, in even the best American colleges, almost certainly not over .80, which means that the official ratings are six-tenths as erroneous as would be the case if the grades were assigned at random by a child, as in a lottery! If 900 students pass and 100 fail by the official ratings in a subject, there is every reason to believe that nearly half of those who failed really did better than some of those who passed. . . .

It surely is unwise to give instruction to students in disregard of their capacities to profit by it, if by enough ingenuity and experimentation we can secure tests which measure their capacities beforehand. . . .

On the whole, it appears that the effort to replace opinion by measurement in our ratings of the achievement of higher education will increase and spread rapidly. Indeed, it may soon need protection from over-extravagant hopes more than from hostile criticism.¹

THE NEW TYPE TESTS

Out of the criticisms of the traditional examinations on the ground that the marking is subjective and consequently neither valid nor reliable, of standardized tests for the reason that they set up norms which are not universally applicable, and of intelligence tests because they are neither specific nor comprehensive in what they measure, there arose the new type test which, while made possible by the techniques already established for the development of standardized and intelligence tests, may be used to test informally any subject of the curriculum at any level of instruction.

¹ E. L. Thorndike, "Measurement in Education," National Society for the Study of Education, *Twenty-first Yearbook*, pp. 1-8 (Bloomington, Ill., 1922). Quoted by permission of the Society

The theory upon which the measurement of educational products is based has been clearly defined by Professor Thorndike:

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality. Education is concerned with changes in human beings; a change is a difference between two conditions; each of these conditions is known to us only by the products produced by it—things made, words spoken, acts performed and the like. To measure any one of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. To measure a product well means so to define its amount that competent persons will know how large it is, with some precision, and that this knowledge may be competently recorded and used. . . .

In general the nature of educational measurements is the same as that of all scientific measurements. In detail, however, there are notable differences. An educational product, such as a composition written, a solution of a problem in arithmetic, an answer to a question, or the performance of an errand, is commonly a complex of many sorts of things.¹

The new or objective type of examination is a form of test which seeks to reduce the weaknesses and defects of the written or essay type of examination and which aims to yield results which are objective and more valid, reliable, accurate, and comparable in similar situations than the old. The chief characteristic of the new type test is that it calls for a short answer by one or a few words, by a check mark, or by a number, and can be answered only if the individual under examination has the pertinent knowledge immediately and readily available. A very marked and important difference is that, while in the essay type of examinations the number of questions to be answered is limited particularly by the amount of time required for writing the answer, the new type test provides for a wider sampling of the content of the subject of the examination and is more comprehensive both in scope and extent and in range of difficulty. In place of the small number of questions which may be asked in the essay or long answer test, as many as from 50 to 100 or more items may be answered in an hour's new type or short answer examination. From this point of view alone the new type test reduces the element of chance or luck so far as the examinee is concerned because of the comprehensive range of the questions selected; to this extent it is fairer to the students.

A further advantage claimed for the new type test is that since the answers

¹ E. L. Thorndike, "The Nature, Purposes, and General Methods of Educational Products." National Society for the Study of Education, *Seventeenth Yearbook*, pp. 16 and 17 (Bloomington, Ill., 1918). Quoted by permission of the Society.

must be short and specific the irrelevant factors, which according to investigations may distract attention from the real marking of the essentials, are removed or reduced to a minimum. This means in turn that the personal equation or subjective opinion does not affect the judgment of the examiner. The marking or scoring is thus objective on a scale which consists of a series of small units easily agreed upon by competent examiners. The results are more accurate because only one variable, whether the answer is correct or incorrect, is measured at a time.

It is important to note a still further distinction between the traditional and new types of examination. The preparation of questions which call for a long answer does not on the whole take a great deal of time, while the correction of the papers is an involved and laborious process. The new type test, on the other hand, is more easily administered and marked, but its construction requires far more time and care, a fact which is frequently ignored. To a far greater degree than the essay type the new type test arises more directly out of the materials of instruction and represents the results of an extensive survey of the subject examined on the basis of courses of study, textbooks used, and the opinions of experts. The construction of the new type test brings together in a coöperative effort the teachers of a subject and the experts in measurement. The new type test may, however, have the same character of detachment as any other form of external examination. For this reason it is subjected to the criticisms that it controls the organization of the content of courses of study, that it leads to special preparation, that it interferes with the flexibility and adaptation of courses of study.¹

It is doubtful whether these criticisms are warranted; they are a repetition of the traditional criticisms of examinations and come in the main from that school of educational thought which claims, on the one hand, that any form of test tends to warp the main functions of education—the promotion of growth, the development of personality, and training in “how to think rather than what to think”—and, on the other, that the same results can be attained by the progressive methods as are achieved by the traditional. The practice of measurement is based on nothing more than the premise that if a subject is studied the school is justified in assuming that there has been some achievement and that that achievement should be measured in the interests of the orderly educational and instructional progress of the pupils. From this angle any form of test which attaches importance to the organization of curriculum according to subjects, which lays stress on content values,

¹ See Selections 880–886 in W. H. Kilpatrick, *Source Book in Philosophy of Education*, pp. 455 ff. (New York, 1934).

and on systematized knowledge, facts, and information is open to objection from those who wish to organize the work of the school on what it termed an activity program.¹

More fundamental and relevant is the criticism, particularly from those who rightly value the merits of the essay examination and fear its disappearance, that the new type test, consisting of 100 or more items to be answered in a short time, is a test only of memory of facts, information, or knowledge, and that it cannot serve as a test of reasoning or the logical processes of thought as may be indicated in the organization both of knowledge and ideas called for in an essay examination. This criticism has not been ignored; consideration and discussion of it have gone far enough to prove that it is not valid unless it is assumed that the recalling of facts and information and thinking are independent functions of intelligence, and that reasoning is largely dependent upon the external form as opposed to the substance of a question. The first question suggested by this criticism is, What is the place of facts, information, and knowledge, and what is their importance in the equipment of the educated man as well as the professional worker? The question has been well answered by President A. Lawrence Lowell:

Throughout our lives we are constantly forced to muster all we can of our previous knowledge, and the habit of doing so can be cultivated by practice. How often when the occasion has passed do we ask ourselves, as a student does after an examination, why we did not remember some essential fact. The art of recalling quickly, fully, and accurately is certainly a valuable part of mental training. It is a special art, not the same thing as a rich store of knowledge. Some men have all the knowledge they possess ready for use on demand; some require a certain time for reflection before they can produce it; and some can make use of it only in the solitude of their studies.²

Actually it is impossible to make a clear distinction between thinking and memory of facts by means of a test.³ To the criticism that the new type tests are only tests of memory, the answer given by Professor Ben D. Wood in

¹ On the question whether progressive schools should abstain from the use of tests see R. W. Tyler, "Evaluation: A Challenge and an Opportunity to Progressive Education," *Educational Record*, January, 1935, pp. 121 ff. "Experience shows that many of the areas which are now called intangible and incapable of evaluation can be appraised and that this appraisal can become more and more accurate and valuable" (*Ibid.*, p. 128.)

² A. Lawrence Lowell, "The Art of Examination," *Atlantic Monthly*, Vol. CXXXVII, January, 1926, p. 62.

³ See the study previously mentioned by R. B. Weaver and A. E. Traxler, "Essay Examinations and Objective Tests in United States History in the Junior High School," *School Review*, Vol. 39, 1931, pp. 689 ff.

1923, when the new type test was in its infancy, has been confirmed both by subsequent experience and by experimental studies.

There is not as much opposition between "information" and "reasoning" as some teachers would have us believe. Facts do not exist in the mind in isolation. We remember by thinking and we think by remembering facts. Those who declaim against tests of "mere facts acquired and remembered," should realize that we cannot think without facts. We could not neglect the measurement of facts without neglecting the measurement of the very fabric of thinking. When we consider that facts are not only a legitimate and undoubted aspect of thinking, and also that they can be acquired, retained, and reproduced only by thinking, only by organizing material in a logical and systematic manner, there can be no doubt of the value of a pure information test. In fact, the measurement of reasoning and organizing ability in a field of knowledge would be very defective and incomplete without a measurement of the breadth of information in that field. Breadth of information cannot be tested with the essay form of examination in a three hour period, for the simple reason that on physical grounds only so many pages of prose may be produced in three hours; and even if this defect were absent, range of information could not be objectively and reliably measured with the essay form of examination.

Fortunately, we do not have to depend on theoretical considerations. Every experimental study thus far made and reported has shown a very high relationship between measurements of information in a field and intelligence or ability to think in the material of that field.¹

Those who are inclined to criticize the new type test on the ground that it only tests memory and not the "higher processes" of thinking either minimize the importance of a knowledge of facts and information as the necessary material for thinking or rely on a psychology which cannot be defended. One may have the capacity to think but that capacity is idle without a thorough and comprehensive mastery and retention of knowledge and ideas which can only grow out of knowledge. The case for the new type test has been excellently presented by Dr. William S. Learned:

How will it profit us even in case we do secure a trustworthy index of a student's knowledge? In short, just how indispensable is knowledge in education?

There appears to be no way of evading the primary truth that thinking, which is assumed to be the fundamental process in education, if done at all, must be done by ideas or concepts which the mind presents on demand of

¹ Ben D. Wood, *Measurement in Higher Education*, pp. 162 f. (New York, 1923). See also Ben D. Wood and F. S. Beers, "Knowledge versus Thinking" *Teachers College Record*, 1936, pp. 487 ff.

any given need or situation. Furthermore, the success or satisfaction resulting from thought will clearly be in such proportion as the concepts thus provided are adequate both in number and in clearness of definition. A "resourceful" mind is one which, at each step of its procedure, is supplied automatically with such a variety of clear-cut, relevant suggestions that its problem becomes one of selection. Several possibilities are normally present—and present in such form that a choice can be made intelligently because each possibility assumes to the main situation relations that are sharply defined. The ineffective mind, on the contrary, offers little or nothing to meet a need; its proposals are either hearsay recipes that have no real application or else are so vague and cloudy as to be worthless.

The test of education is therefore plain. Its business is, first, to increase the material which the mind thus has ready for use. To this end it seeks to expose the student to important facts, ideas, and values as widely as these can be effectively appreciated and absorbed. Second, and far more important, it must ensure the availability of this store of material by progressively refining and clarifying its *meanings*. Through frequent restudy and redefinition of the ideas themselves, by approaching them deliberately and repeatedly from fresh points of view, an understanding of their bearings and interrelationships gradually matures and becomes an increasingly vivid possession of the student.¹

The criticism is based, furthermore, on an assumed requirement that all the steps in a process of reasoning should be presented, as is expected in an essay, as well as the end-product, without realizing that this end-product must have been reached by a process of reasoning even though the steps are not written out in full.

As for the argument that the essay examination measures thinking, while the objective test measures *mere* memory, more heat than light is generated by its advocates. The proponents assume, usually unconsciously, that a student must present the steps in his reasoning before they will accept the end-product as more than mere association. It is no weakness in the objective test that it demands of the student that he shall do his thinking subvocally and without movement of the wrist in writing before indicating the end-product of his thinking.²

It cannot be repeated too often, however, that those who have been the leaders in advocating the use of the new type test have not suggested that it should displace the essay as a form of training and practice in the clear organ-

¹ W. S. Learned, "Testing for Values in Education," pp. 2 f., in Reprint from the *Bulletin of the Association of American Colleges*, Vol. XX, No. 1, March, 1934.

² M. S. MacLean, "The Problem in the General College," in *Studies in College Examinations*, p. 14 (University of Minnesota, Minneapolis, 1934).

ization and logical presentation of facts, concepts, and ideas. The only claim that they make is that the new type test is a better instrument not only for measuring the amount of information acquired and reasoning ability, but also for other purposes which are important for educational controls. The possibility of formulating essay questions in such a way that the answers can be marked objectively and yield more reliable results than marking by general impression is not excluded, but the more accurately and carefully this is done, the closer will be the approximation to the new type test and the less in the long run will be the real educational value of the essay. In a recent study of the two methods of examination the results were summarized as follows:

Research with the essay test has, in the main, been conducted with the traditional or unimproved form and comparisons have been made with improved forms of objective tests. Such comparisons indicate that the improved objective test measures as accurately as or better than the unimproved essay test. When analysis of essay tests into defined types such as *list, outline, describe, compare, contrast, explain, discuss, develop, evaluate, and summarize* is achieved for instructional use, comparative studies using improved essay types and improved objective types of tests will become possible.

The use of both objective and essay tests seems to be a better basis to evaluate the achievement of such varied results as specific, general, and organized information.¹

The new type test is considered superior to the old for a variety of reasons, among them the following:

(1) It more nearly meets the standards of definable validity, that is, it measures to a known degree what it is designed to measure to the exclusion of factors irrelevant to this particular end.

(2) It is statistically more reliable in the sense that the scores on two equivalent new type tests given to the same pupils agree more closely than do the scores on two supposedly equivalent essay type tests of equal length.

(3) It is more objective since subjective factors such as bias, prejudice, personal opinions, and temperament are reduced to a minimum and the scores are so accurate and consistent that they would always be the same no matter what the number of examiners might be.

(4) It is accurate in the sense that the units of measurement are more nearly equal at all points of the scale.

¹ C. C. Weidemann, "Recent Developments in the Written Essay Examination," *Review of Educational Research*, Vol. V, December, 1935, p. 489. A list of investigations on the subject is given on pp. 516 f. of this issue.

(5) It is more comprehensive because it includes a wider sampling of the subject of the examination and to that extent is also more reliable.

(6) It is easily administered and marked with economy of time and effort for both examiners and pupils, but it requires more time to prepare than the traditional examination so that the standards of scientific measurement may be adequately met

(7) The results can be interpreted more easily and used for specific purposes of improving weaknesses discovered in pupils or encouraging ability.

Hence the new type test by new methods and techniques seeks to meet those difficulties which are encountered in the attempts made to secure accuracy and reliability in the marking of the traditional examinations by consultation, conference, supervision, and rechecking but, as experiments have proved, without great success. Unlike the essay type in external examinations the items in a new type test are based on matter actually covered or assumed to have been covered by a given group in a given course and are based on courses of study, textbooks, and the opinions of subject matter experts. They can thus probe into every phase of a subject in order to secure a wide sampling of knowledge and to this extent eliminate from instruction the attempts to guess at or to emphasize or gamble on "examinable" topics.

The new type test should accordingly not interfere with the process of instruction as is so often the case when external examinations are the object of preparation. The construction of the new type test involves far more time and care than the preparation of the traditional examination. As a rule more questions are prepared than will be used in one test in order to afford a basis for the selection of the best. Questions which are ambiguous as to meaning and possible answers are excluded. The items are short and so selected that even the poorest student can answer some correctly and the best the majority but not necessarily all; a series of items may accordingly be arranged in equal numbers of easy, moderately difficult, and difficult character. The students are given specific directions usually accompanied by an example. The test is as a rule "timed." The answers are written in alongside of the questions and can be easily scored with a key, so that expertness in preparing the test is substituted for expertness in correcting the answers.

A further advantage claimed for the new type test is that it lends itself to organization in a variety of forms, while the essay type of question is, as a rule, restricted to a few limited forms, such as "describe," "discuss," "compare," "give reasons for," "explain," and "tell what you know about." The most usual forms of new type test which have been developed and used are the following:

(1) Recall type which requires the insertion of a word to make a statement true.

(2) Somewhat similar to the recall type is the completion test in which missing words are to be supplied in a series of statements.

(3) Recognition type, in which a number of suggested answers to questions are supplied and the student is required to select the best or the correct answer from among them. The recognition type of test may take one of three forms: (a) the True-False test in which the student has to judge the truth or falsity of a number of statements; the answer may be given by underlining the words True or False, or the letter T or F, or by Yes or No, or by a plus or zero sign. From such evidence as is available the objection to this test that a student may learn a false statement is not warranted. On the other hand, it is obvious that the true-false test does not lend itself equally well for use in all subjects. That the element of guessing may enter is generally admitted; its influence is, however, reduced by the use of the formula Rights minus Wrongs in determining the score. (b) The Multiple-Choice Test or Multiple-Response Test which involves the choice of a word or phrase which makes the best answer in a statement. The answer may be given by underlining the appropriate word, or by inserting a number, if the suggested answers are numbered, or by putting a check against the best answer. (c) The Matching or Association Test involves the matching or pairing of two sets of associated facts each of which are arranged in parallel columns in chance order.

(4) Other forms of tests are in general variations of these more common forms, such as the Analogies Test, Similarities Test, a Classification or Cross-out Test. Each of these tests may be used independently or in combinations to form a "battery of tests." The simplicity of the finished product is likely to be delusive since more is involved in the preparation of test items than inverting or disguising questions. The amateur is likely to plunge into the preparation of tests without realizing the editorial and statistical difficulties involved in constructing and scaling them, in interpreting scores, or in converting scores into grades.¹

PURPOSES OF EXAMINATIONS

The purpose of the traditional examinations was in the main to discriminate between students who should pass or fail on the basis of some precon-

¹ The following brief list of references will serve as an introduction to the subject:

A. R. LANG, *Modern Methods in Written Examinations* (Boston, 1930).

W. A. MCCALL, *How to Measure in Education* (New York, 1922)

W. S. MONROF, *Written Examinations and their Improvement* University of Illinois Bulletin, Volume

ceived standard of achievement. In general the results of such examinations have been as immutable as the laws of the Medes and Persians. To a certain extent they helped to standardize the instruction of the schools or educational institutions which were subjected to them; to that extent also they have often been used as much as a test of teachers as of their pupils, and the quality of a school has been judged by the number of its successful pupils. Underlying this practice have been the assumptions that the standards of assessment were always unchanging, that the opinions and judgments of examiners were invariably infallible, and that the curriculum and courses of study were always adjusted to the needs and abilities of all pupils or students at a given level; hence the pupil or student who failed stood condemned as incapable and was separated by a wide gap from the one who just managed to pass. Examinations of such a type have rarely been used either as a means of discovering where a pupil's weaknesses lie or as a basis for remedying these defects; still less have they been employed for diagnostic purposes, that is, to discover the type of course from which a pupil is most capable of profiting. The traditional examination, particularly if conducted by an external board, allowing for the somewhat greater choice of subjects which has been permitted, is still built on the assumption of identity of abilities rather than on a wide range of variability. Hence the difference of a few marks may determine whether a student is to be admitted, in a European country, for example, to a secondary school, or to a higher institution. Too frequently a career depends upon the hazard of one examination, marked by examiners who in order to maintain standards of fairness, detachment, and impartiality prefer to know nothing more about the candidate than is revealed in his papers.

Such a conception of examinations may have been justified at a time when opportunities for secondary and higher education were selective on some basis or other, although there is ample evidence of dissatisfaction with and criticism of the traditional examinations for more than half a century and not always for technical reasons alone. The situation, however, is changing everywhere. The task which confronts educators to-day is not one of separating the sheep from the goats, of dividing the population into those who have passed and those who have failed, or of setting up the curriculum as a hurdle

20, No. 7, *Bureau of Educational Research Bulletin*, No. 9 (Urbana, 1922).

C. W. ODELL, *Traditional Examinations and New-Type Tests* (New York, 1928).

D. G. PATERSON, *Preparation and Use of New-Type Examinations* (Yonkers, N. Y., 1925).

G. M. RUCH, *The Improvement of the Written Examination* (Chicago, 1921).

———, *The Objective or New Type Examination* (Chicago, 1929).

C. RUSSELL, *Classroom Tests* (Boston, 1926).

P. M. SYMONDS, *Measurement in Secondary Education* (New York, 1927).

BEN D. WOOD, *Measurement in Higher Education* (Yonkers, N. Y., 1923).

to be overcome. With the inevitable prolongation of "infancy," in the sense used by John Fiske, due in part to economic reasons and in part to the extension of the compulsory age for school attendance, the problem has ceased to be one of selection and has become one of the distribution of education, or the discovery of "the right education for the right pupil under the right teacher." The success of an educational system can or should no longer be measured in terms of the numbers who pass or fail in examinations but by the degree to which it has been able to discover the abilities and needs of pupils and students and has provided for them the type of education from which they are capable of profiting. And for this purpose the traditional type of examination cannot be used, for the problem is not merely one of selecting an *élite* or even *élites*, since the types of leaders required by modern societies have increased in numbers, but of giving to all the type of education and instruction that will equip them, in accordance with their abilities, to be useful members of society.

The question here involved may be asked in another way—How much education has an individual received if he leaves school with a record of failure or with a record of just having passed, and so on? Those who fear to approach the question in these ways are alarmed, first, lest standards will inevitably be lowered, if more pupils or students pass or, second, lest the introduction of modifications in the established concepts of educational curricula lead to a dilution of the best education for the best minds. The recent educational history of France and Germany furnishes ample evidence of the fact that the large increase in the numbers attempting to meet the *existing* requirements of secondary education has provided sufficient justification for the first fear. On the other hand, if society through its educators undertakes to discover types of education best suited to the great variability of intellectual capacity the second danger need not be imminent. It was a recognition of the possibilities of this solution that prompted those who have been leading in the movement for the creation of the *école unique* to recommend the creation of a National Bureau of Selection and Guidance. No steps had, however, been taken to devise either methods or curricula to implement this suggestion.¹

If the traditional examination method has ever been used for any of these purposes, it has on the whole been used in a negative way, to weed out those who were not likely to succeed in a particular type of course or career rather

¹ That the French *baccalauréat* created for one purpose may no longer be adapted to meet the demands of changed social and economic conditions is suggested in a discussion of this examination in the French Committee's *La Correction des Epreuves Ecrites dans les Examens*, pp. 28 ff. (Paris, 1936)

than to discover the nature of abilities and to make the necessary adjustments to them. The new type test serves most of the purposes which the traditional examination may in the minds of some have been supposed to aim at but which it fulfilled only empirically and in a snapshot rather than a detailed manner. The new type test furnishes more reliable information as to a pupil's achievement and, because of the range of the samplings, provides a better clue to his weaknesses; in the case of the able pupil evidence may be secured as to whether he is working up to his capacity. It is thus an excellent basis for educational guidance, and, indeed, this function is beginning to be regarded as more important than the examination itself.

The first question that the school should ask and answer at least provisionally several times each year, writes Professor Wood, is "What *can* Johnny learn, and which of the things he can learn *should* the school, in the light of all the facts, try to help him to learn? . . . Tests should first of all tell us *what* the pupil should *try* to learn—not *how* he may be cajoled, persuaded, or insidiously coerced into learning item *x* in the standard curriculum for grade *n*. If a pupil has difficulty in learning item *x*, this fact in itself may be evidence that *x* is not suited to his capacity and needs and that he should, therefore, be given opportunity to learn something else, not forced by remedial treatment to live up to the "high and ever higher standards" so often found in perorations that are more impressive than meaningful. . . . Pupils who after a year or two of trial cannot learn might be excused even from some of the supposed essentials of the elementary curriculum and encouraged to study some vocational subject or subjects that lie within their capacities and interests and thus have meaning for them. In many cases this procedure would not lessen their achievements in the elementary curriculum, and might engender better habits and more responsible attitudes in place of the disappointments that now frequently result.¹

This idea Professor Wood repeats in another form as follows:

The duty of educational institutions of non-professional character is to help boys and girls to become better citizens, with the aid of the curriculum, if possible, but in spite of the curriculum, if necessary; and that the method of achieving this objective is to mobilize half our educational resources, if necessary, to the task of ascertaining the abilities and needs of pupils as growing individuals, and to use our remaining resources to meet those needs, whatever those needs happen to be.²

¹ Ben D. Wood, *Basic Considerations in Educational Testing*, pp. 9 ff. (1933).

² Ben D. Wood, "The Ultimate Basis for Satisfactory College-High School Relations." *Bulletin of the American Association of Collegiate Registrars*, July, 1934, p. 273.

Introduced and developed first in order to secure more valid, accurate, reliable and objective results than were obtainable by the old type of examination, the new type test has been welcomed as a valuable instrument for educational guidance, as the following statement by Dr. C. R. Mann, formerly Director of the American Council on Education, indicates:

Along with more skillful production of such measuring instruments has come increasing recognition of their value. Now the testers tell us that their tests are proving more useful as instruments for educational guidance than for ranking students in the order of their proficiency in particular school subjects. They reveal what sort of things each student does well as well as the things he does badly. Emphasis can then be given in each student's program on things he does well rather than on those he does badly. When this is done, he will be more profitably employed in school than if he is driven through a set curriculum of traditional subjects many of which he does not get at all.¹

The value of the new type test would even be increased by frequent repetition, provided accurate records were kept and suitably interpreted.

We should rather multiply examinations and tests, of many kinds, using them frequently, *but always informally, casually, and skeptically*; record the results, of course; correlate and study these results; study particularly the patterns of results in each student's cumulative record, in conjunction with personal impressions, teachers' grades, or, better, teachers' estimates, and all available facts in regard to the student's background and achievement; and base the necessary administrative decisions, with respect to graduation, promotion, classification, and guidance on *the total picture* of the student's abilities, aptitudes, character and potentialities—to which total picture a considerable number of comparable test results would seem to me to contribute a vitally necessary part.²

This function of the new type test aptly fits in with recent tendencies in the concept of educational administration and organization, for, according to Dean H. E. Hawkes of Columbia University,

in order to do our jobs either in schools or in our colleges, we need and must have more information than has ever been available concerning the characteristics, capacities and past achievements of our students, if we are to do them justice or hope to approach our own standards of accomplishment.³

¹ C. R. Mann, "The N.R.A. and the Schools," *Educational Record*, October, 1933, p. 452

² Max McConn, "Measurement in Educational Experimentation," *Educational Record*, January, 1934, p. 118.

³ H. E. Hawkes, "Dangers in the Testing Movement," *Educational Record*, January, 1935, p. 32

Finally, the use of examinations is determined not only by a philosophy of education but also by a social philosophy, which has been well stated from the American point of view by the University Committee on Educational Research of the University of Minnesota:

The expanding program of examination activities . . . has, therefore, not taken place without deep underlying causes in the social order. The motivation is primarily the search for quality in human beings. It arises from economic pressures, the surplus of workers, and the competition for effective service. In the schools, the search is for quality of mind, chiefly intellectual quality, that can be trained by the processes of education to higher competence. Although the evidence that examinations are assuming a new rôle in American life is unmistakable, it is still unlikely that we shall surrender our national philosophy of equal educational opportunity for all the children of all the people. At every advancing step American society will insist that examinations shall be of the sort that opens the doors of opportunity rather than closes them. They must reveal what young people can do and not merely what they cannot do. They must not only be fair and accurate, but they must canvass the resources of the individual as well as his limitations. The American outlook will not long abide a merely negative program, but will insist on one which reveals possibilities even though it closes innumerable doors which seemingly have hitherto stood open.¹

The new type test lends itself better than the traditional examination to improving the quality of instruction, because the points in which pupils are strong or weak are more clearly indicated in detail. It may be used for the promotion of pupils from class to class or school to school, and it may be employed further for comparing the results attained in different classes, different schools, and different school systems. It is an instrument which, if wisely used, can serve a number of educational and instructional purposes, but as in the case of standardized tests and intelligence tests, it has not met the expectations of those who are still looking for a single instrument which will measure all the purposes of education—intelligence, personality, attitudes, character, vocational capacity, achievement, and promise of success in school, college, or after-careers.²

The main uses of the new type test in elementary and high schools is to

¹ The *Bulletin* of the University of Minnesota, University Committee on Educational Research, *College Examinations*, Vol. XXXV, No. 22, April 11, 1932 (Minneapolis, Minn., 1932).

² See National Society for the Study of Education, *Twenty-first Yearbook*; P. Symonds, *Measurement in Secondary Education*, pp. 1 ff. (New York, 1927); P. Mort, and A. I. Gates, *The Acceptable Uses of Achievement Tests* (New York, 1932); and *Review of Educational Research*, Vol. V, No. 5, December, 1935, pp. 445 ff. and pp. 500 ff.

discover the ability and achievement of pupils and to advise and guide them appropriately. Since the high school is open to all who complete the six or eight grades of the elementary school without an entrance examination the function of the test becomes increasingly one of securing an adequate basis for guiding the pupils in the vast maze of choices open to them in the modern secondary school curriculum. The test accordingly serves a dual purpose: measuring achievement and at the same time providing one basis for predicting success; or the two purposes may be described as diagnostic and prognostic. In general, the school is as much concerned with the discovery of causes of failure as it is with the prospects of success. This means in turn the provision of such varied curricula and courses of study as will make possible adaptations to pupil capacity and interests. A recent analysis of the "Present Tendencies in the Uses of Educational Measurements" closes with the following summary:

By way of conclusion, it may be said that the summary of the previous pages bearing upon the present tendencies in the uses of educational measurements indicates clearly that tests and measurements for their own sake are rapidly passing from the educational picture. On the other hand the use of tests as an integral and essential part of educational procedure and research is growing. Using tests for what they may contribute to the realization of the important aims of education and the solution of educational problems appears decidedly to be the modern tendency ¹

INCREASING ENROLLMENTS AND THE USE OF TESTS

Interest in the development of adequate instruments for testing and measuring the abilities of students in such a way as to secure valid, reliable, and comparable results was increased by two important factors: First, the rapid increase in enrollments in the high schools and colleges of the country after the World War and more particularly in the depression period following 1929; and, second, an attempt to implement and redefine the traditional ideal of equality of educational opportunity by discovering the right education for the right student. Both factors combined to lead to the acceptance of the principle that the emphasis in educational administration and organization should be placed not upon the selection of the few on the basis of one crucial examination—a snapshot view, as it has been called—of the abilities of a

¹ *Review of Educational Research*, Vol. V, December, 1935, p. 468. See also G. Hildreth, "Tests of Mental Capacity—High School Level" in *A Bibliography of Mental Tests and Rating Scales*, pp. 36 ff., and "Achievement and Educational Tests," *ibid.*, pp. 75 ff., R. Strang, *Personal Development and Guidance in College and Secondary School*, Chs. III and IV (New York, 1934); and P. Symonds, *Measurement in Secondary Education*, Chs. XVII, XIX, and XXI (New York, 1927).

student but upon the distribution of the right education based upon a thorough understanding of the whole individual derived from a study of his abilities, capacities, interests, and needs continued over a period of years. From one point of view this movement has meant a search for an instrument or a series of instruments which would enable educational authorities to guide students in the educational careers most appropriate to their abilities; from another it has intensified the criticisms which have accumulated during the last quarter of a century against the traditional notion that all students must be poured into the same mould of what have been defined as the essentials of a liberal education.

As already pointed out the problem under consideration is an internal one so far as elementary and high schools are concerned because of the absence of any obstacles to advancement from one to the other. The problem becomes particularly crucial at the stage when the admission of students from high school to college arises, for here there are definite entrance requirements in some or most institutions and even under a system of accrediting graduation from high school furnishes no adequate guarantee of a student's ability to succeed in academic work. While intelligence and standardized tests of some kind have been used extensively in elementary and high schools, their use has on the whole been spasmodic and not cumulative in character in the sense of presenting a complete picture of the intellectual abilities and personal and educational records of the pupils.

During and immediately following the War the Army and other intelligence tests began to be used extensively by colleges for purposes of admission. It was soon demonstrated, however, that the relationship between psychological tests and academic success was not very high; in the majority of cases the coefficients of correlation were below .50. The Army Tests were found on the whole too simple for prospective college students, while the other tests attempted by a single measure to evaluate a complex of abilities. Further, since marks or grades in college were awarded on the basis of the traditional examinations and the subjective opinions of the instructors, they were themselves unreliable and therefore unsatisfactory as a criterion of the predictive value of the psychological tests. Despite the low relationship between tests and academic grades the intelligence ratings derived from the tests proved to be valuable to deans and administrative officers in dealing with students and also revealed the presence in colleges of students who under no circumstances could be expected to succeed in the courses actually offered.

EXPERIMENTS WITH NEW TYPE TESTS COLUMBIA COLLEGE

In 1919 Columbia University undertook the experiment of admitting students to the undergraduate College on the results, among other requirements, of the Thorndike College Entrance Intelligence Examination. This Examination was both more comprehensive and more difficult than the Army Alpha Test. The advantage of the Thorndike Examination over other tests was that besides being a test of intelligence it also included educational tests of a type and standard suitable for high school students. Candidates for admission to Columbia College were in all cases required to submit school and character records and might substitute the Thorndike Examination for the usual entrance examinations. The Thorndike Examination has been proved to have a higher predictive value, at least for the first year of college work, than any other measure so far developed. While tests up to the time of the introduction of the Thorndike Examination had in general shown a coefficient of correlation below .50, the correlation between the Thorndike Examination and the freshman year was early shown to be .63, between the Examination and the sophomore year .62, and between the Examination and both years .67. It was further demonstrated that the Thorndike Examination had a predictive value of college success higher than the New York State Regents' Examinations, which was .64, and secondary school marks, which was only .26. The result of this experiment not only at Columbia College but also in other institutions was to shake confidence in the use of the Army Alpha Test which had already begun to be used by a large number of institutions and to be regarded as defective on other grounds.¹

Further experimentation with intelligence tests for admission to college (and it must be noted that the character of the tests themselves had already changed from their original form and sought to test both native ability and educational attainments) has revealed the fact that, while they have some predictive value, they indicate probabilities rather than certainties.

All of the studies agree upon the fact that there is usually a substantial amount of correlation between intelligence test scores and scholarship among any student group. The fact that the correlation is not perfect indicates clearly that factors other than tested intelligence play considerable part in the determination of scholarship levels; but the size of the coeffi-

¹ See National Society for the Study of Education, *Intelligence Tests and their Use*, Ch. X (Bloomington, Ill., 1922); Ben D. Wood, *Measurement in Higher Education*, Chs. II-V (New York, 1923); National Society of College Teachers of Education, *Yearbook XVIII, Quantitative Measurement in Institutions of Higher Learning*, Ch. III (Chicago, 1930); and R. Strang, *Personal Development and Guidance in College and Secondary School*, Chs. III and IV (New York, 1931).

cients of correlation that have been presented leave no doubt that intelligence is one of the basic factors conditioning scholastic success. Efforts to establish a certain point indicative of certain failure or certain success have not been entirely satisfactory, the prediction being yet in terms of probabilities rather than certainties.¹

The tests have indicated the threshold of intelligence or capacity below which success in college cannot be expected, although there is no general agreement as to that threshold.² But while their predictive value has not been perfect their use has played a more important part in directing attention to the necessity of diagnosing individual differences in abilities and interests and in taking into consideration a large number of other factors which may enter into the determination of academic success. Among such factors are health, methods of study, financial status, earnestness, and perseverance, selection of courses, time devoted to extra-curricular and social activities, and so on. In other words, the result of the effort to discover a more reliable method than the traditional examination for selecting students for college has been twofold: it has shown that a higher correlation can be secured between a test like the Thorndike College Entrance Examination and later success in college, and that even such a test cannot be relied upon alone for purposes of prediction. Accordingly attention has been directed to the importance of guidance in education and the desirability of providing cumulative records about each student, covering every aspect of his personality which the adviser should know if guidance and counselling are to be effective. Thus Professor Strang reports that

At Colgate University, during a period of three years, about three-fifths of the failing students were found to have intelligence test scores below the average for the campus. There was a marked difference in the ratio of failing students above average in intelligence to failing students below average in intelligence in the freshman and sophomore classes. It appears that freshman students below average in intelligence withdraw from college leaving students in the upper classes who are more likely to fail because of emotional or other factors. Freeman, using The American Council Psychological Examination, reported a marked tendency for the "students in the higher deciles to achieve better averages, to encounter fewer scholastic difficulties, and to survive a greater number of terms." But scores in the

¹ W. F. Reeves and J. D. Russell, "Some Aspects of Current Efforts to Improve College Instruction." *Bulletin of the Bureau of School Service*, Vol. I, No. 2, December, 1928, p. 14. Published by the University of Kentucky.

² The difficulty is intensified further by the great variety of standards in the colleges themselves, as revealed in the Pennsylvania Study (see pp. 133 ff.).

lower deciles were not clearly indicatives of failure. Nearly half of those in the lowest decile succeeded in completing the full course, though with great difficulty.¹

After considering a number of criteria of prediction which have been attempted, Professor Strang concludes that

Predictions based on the data from cumulative records, involving as they do the many-sided aspects of the individual seem to be the most satisfactory basis on which to advise boys and girls concerning their further educational plans. The cumulative record prevents a single non-typical performance from being given undue weight. None of the criteria can predict, with the certainty that the term implies, an individual's success in college. It must not be assumed, however, that the fault lies wholly in the criteria. The unreliability of college marks and the inadequacy of college courses are responsible in large measure for the imperfect results.²

The search for criteria of predictability resulted in the recognition that no single measure could be perfect and that the problem merges into the general problem of guidance which in turn raises the whole question of the meaning and purpose of education. It may, indeed, be possible to define the threshold of intelligence required for success in college work, to predict in general that one student is likely to succeed and another to fail. This, however, involves the assumption that the present arrangement of the college curriculum is adequate to meet all types of abilities and aptitudes. The lack of a perfect measure of predictability of college success, with the recognition of the importance of guidance on the principle that the function of non-professional educational institutions, especially publicly supported schools and colleges, is to discover the abilities and interests of students and to direct them into channels of study from which they can best profit, has led to the application of the new examination techniques to wider uses. It is not enough to discover that a student is likely to succeed in college; there is the further task of discovering the types of courses in which a student is most likely to succeed. There has accordingly been developed and used in many colleges a series of placement tests whose purpose it is to discover the actual achievements and aptitudes of students in the various branches of a liberal or general education—English, foreign languages, mathematics, sciences, and the social studies. The placement tests, of the new objective type, follow the principles and varieties of this type of examination which have already been indicated.

¹ Ruth Strang, *Personal Development and Guidance in College and Secondary School*, p. 88 (New York, 1934).

² R. Strang, *op. cit.*, p. 133.

The student adviser uses the results of both the intelligence and placement tests as the basis for recommending the choice of programs of study. At Columbia College the Thorndike College Entrance Intelligence Examination and a variety of placement tests are used in advising students:

Freshmen are sectioned, on the basis of the Thorndike examination, in their course in Contemporary Civilization, and in other required courses, in from three to five sections, the placement tests being used for initial guidance in making detailed adjustments. It is the policy of the College to treat each individual case on its own merits and to consider the most comprehensive and accurate information obtainable concerning each student before making a decision. Cumulative records of comparable information covering the student's early years in school are lacking. The record of early years in school would be of great assistance to the adviser, especially since the average reliability of the tests as given at Columbia is only about 0.80, and since tests merely serve to give present status and not the dominant trends of development and interests of a student nor the underlying history of a case.

At Columbia tests are used not only for placement but for continuous replacement throughout the student's entire work at College, whenever it seems desirable. Tests are repeated or new tests are given throughout the first and second years. In the fall of 1928, the placement tests resulted in the saving of an aggregate of about 1700 units of college work by promoting students in English, foreign languages, and mathematics beyond their official entrance status. On the other hand, a considerable number of students were demoted below their entrance credit status because of deficiencies shown in the placement tests. Students are encouraged to consider their placement as tentative and every student has the constitutional right of petition to raise or lower his placement at any time during his college career. There is thus built up for each student a mass of centralized and correlated information, with comparable tests related to local standards based on several years of accumulated data.¹

The shift in emphasis in the use of tests is brought out in the following summary of the status of Achievement Tests in Colleges and Universities:

The studies during the past three years have extended the uses of achievement tests in college. Formerly, their uses were usually restricted to marking students. More and more they have been used in selecting and guiding students, in predicting college success, in placing students within courses and within sections of the same course, in assigning advanced college credit on the basis of examination, in awarding scholarships, in studying the

¹ A. S. Raubenheimer and F. C. Touton, "Present Status and Use of Objective Tests in Institutions of Higher Learning," *National Society of College Teachers, Yearbook XVIII*, 1930, p. 79.

characteristics of college students, in granting diplomas or honors, in studying the effectiveness of educational procedures, and in evaluating teaching ability. The variety of uses is a chief reason for the barrage of criticism which has been directed against many of the tests constructed during the past ten years. The commonly accepted technics of test construction do not produce tests appropriate for such varied usage.¹

One result of the experimentation in the field of examinations which has been conducted during the past fifteen years deserves mention. The search for a measure of prediction failed to some extent because of the unreliability of instructors' marks which were given as a rule on the written or essay form of examination; that is to say, the attempt to correlate results obtained on an objective basis of new type tests and grades or marks assigned on the basis of subjective opinion was found to defeat the ends of the new experiments. Here and there college instructors had already begun to experiment in their own classes with new type tests. Their use on a large scale was introduced at Columbia College in connection with a new course on Contemporary Civilization required of all first-year students. Greeting them at first with skepticism instructors soon became convinced of their utility; in some cases perhaps they became too convinced, since the value of the essay examination was in danger of being depreciated, and more particularly by those who had been in the habit of employing it regularly than by the advocates of the new type test. From the experience in the course on Contemporary Civilization the use of the new type test spread to other fields at Columbia College, as, for example, physics, government, zoology, economics, philosophy, history, English, and civil engineering. The development of such new type tests on the content of a course involves the cooperation of those who conduct the course and of the expert in measurement. On the instructional side this means a clear definition of the objectives of the course and of the content which students should be expected to master. Then, in the words of Professor Wood,

The definition once agreed upon, the next step is to construct an examination comprehensive enough to satisfy that definition, objective enough to insure that competent examiners everywhere will secure comparable results, and reliable enough to make the results worthy of being used by administrators, instructors, and research scholars. The examination may be made comprehensive by the use of the New Type questions, whose brief recording processes make it possible to include a comparatively large number of each of several different kinds of questions. The brevity of the

¹ R. W. Tyler and F. P. Frutchey, "Achievement Tests in Colleges and Universities" *Review of Educational Research*, Vol. V, December, 1935, p. 493.

answers to the New Type questions not only rescues a very large part of the examination period from irrelevant activities and considerations on the part of the examiners, but at the same time renders the examination objective and reliable.¹

The experiment undertaken at Columbia College is not merely an innovation in the method of constructing examinations but represents also a break with the American practice in which each instructor is usually responsible for the preparation and marking of examinations in his own class. In a sense it introduces the principle of external examinations, except to the extent that instructors do cooperate with the technical expert in their preparation, and sets up common objectives for similar courses and standards which are comparable from year to year. It introduces also another principle, that the function of the examination is not merely to encourage and stimulate students to review their work and to mark and rank them on the results, but also to discover more about the students' strength and weakness and promise of further growth, and to advise them on the basis of the information obtained. In other words, the tendency is to look upon the examination itself not merely as a method of assessing achievement but as a method of education.

EXPERIMENTS WITH NEW TYPE TESTS: UNIVERSITY OF MINNESOTA

The whole problem of examinations has been attacked from this point of view at the University of Minnesota in general but more particularly in the recently established General College, which will be described more fully later as an experiment in educational adjustment. As Professor M. E. Haggerty, Dean of the College of Education, has stated, "It is still far from clear whether the trend toward increased emphasis upon examinations will, in the long run, improve the quality of education. Most certainly it will not do so unless the quality of examinations is improved," so that they "may safely be used to measure human traits and achievement, and evaluate the merit of educational progress." Hitherto attention has been devoted to evaluating the final result. "Educational research," in the opinion of one of the instructors, "is rapidly forcing us to recognize that the construction and administration of examinations is a major part of the teacher's job."

The General College has accordingly adopted and developed the practice of giving comprehensive examinations of the new type in all subjects taken by the students. The construction of the tests involves the cooperation of the

¹ Ben D. Wood, *Measurement in Higher Education*, p. 282 (New York, 1923). See also Chs VIII-XII of this work in which the new methods of content examining in various subjects are discussed.

subject matter instructor and the counsellor who is an expert in the preparation of tests and has a general knowledge of the content of the course to be examined. His function is to suggest "the types of test items useful in measuring particular kinds of content or the achievement of particular objectives"; he guides the assistant in the course "by helping to devise items, by criticizing and revising the items prepared by the assistant, and by directing the organization of the tests and examinations"; and he makes or directs "studies of the tests after their administration." The assistant acts as a liaison between the subject instructor and the counsellor, follows the work of a course, studies the materials of instruction, and makes the technical studies under the direction of the counsellor. The instructor determines the teaching objectives and content of the course, acts as the final judge of the accuracy of the key statements in the tests, and approves the inclusion or exclusion of items in the test. The types of tests used include the essay, true-false, modified true-false, multiple choice, reverse multiple choice, multiple response, matching, master-list, classification, completion, controlled completion, and arranging items. The purposes which the General College has set before itself in developing the system of comprehensive examinations are as follows:

- (1) As in course examinations, the comprehensive examinations should measure the several outcomes of instruction. Not only retention of factual materials and applications taught in class, but also ability to make practical applications of principles to new situations and attitudes toward controversial subjects should be measured. . . .
- (2) The comprehensive examination in a given field should extend to materials covering a wider range of information than is included in a course of instruction. . . .
- (3) The comprehensive examination should demand integration. Test situations should be devised which will demand the drawing together of information and principles from different phases or divisions of a field.¹

The experiments with the new type test have been continued over a sufficiently long period and under sufficient control to warrant certain conclusions. There is general agreement that the new type test has advantages which the essay examination does not possess: it tests a wider sampling of facts, it is easily and accurately scored, and it is free from "vitiating opinion which masquerades too often as judgment." The essay does not, however, lose its importance as a method for training in power of expression and ability to organize material, and it can further be used as a test of these qualities, and

¹ E. A. Carstater, "The Plan in Operation," in *Studies in College Examinations*, pp. 344. (Minneapolis, 1934)

as an exercise should be dealt with separately. A special study of Essay and Objective Examinations in English Literature concluded that "much that is being measured with essay tests can also be measured with objective examination items, for on both types students obtain similar ratings."¹

On the other hand, it is becoming equally clear that certain educational values cannot be measured by objective tests; among these is "the ability to discuss a topic in the social sciences, clearly, convincingly, and agreeably."²

The investigations also tend to disprove the charge that the new type test measures only facts and information.

The findings point to the inference that the ability to acquire information on the part of the students in these General College courses has been accompanied to a substantial extent by the ability to apply this information. The measurement of the extent to which this association occurs is an inherent part of a valid examination in those courses where the objectives of instruction include the ability to use, as well as to acquire knowledge.³

The same conclusion as this was reached in another study, "Testing the Understanding of Ideas," and is expressed as follows by the investigators:

It has been shown that the testing program in "Background of the Modern World" involved the student's ability (1) to discriminate, (2) to recall or establish sequences, relationships, and associations, and (3) to apply and contrast generalizations. Thus the examinations measured some aspects of complex mental processes. No matter how inclusive a definition of ideas is set up, it seems clear that the test forms described in this article do test mental activities which would fall within the range of such a definition. No one would claim that the forms measure all kinds of ideas. All that is claimed for them is that they measure some of the more advanced forms of mental activity.⁴

Finally, "the chief advantage of improved examinations accrues from the greater possibilities afforded for reliable uses," that is, for diagnosing a student's special abilities and difficulties as a basis for guidance and advice. The methods of scoring and recording the results make possible the development

¹ A. C. Eurich and F. S. Appel, "Essay and Objective Examinations in English Literature," in *Studies in College Examinations*, p. 97.

² A. C. Krey, "Differential Functions in Examinations," *Bulletin of the University of Minnesota*, University Committee on Educational Research, Vol. XXXVI, No. 4, January 25, 1933, p. 12.

³ P. O. Johnson, "Differential Functions of Examinations," in *Studies in College Examinations*, p. 47.

⁴ E. R. Wesley and R. R. Wesson, "Testing the Understanding of Ideas," in *Studies in College Examinations*, p. 73.

of profile charts which with other relevant information may present a complete picture of the physical, emotional, and intellectual characteristics of a student. Hence

With improved examinations and with the students' ratings shown on profiles the educator and counselor is in a better position to diagnose difficulties that may be ameliorated just as the physician is better equipped to treat a patient the more refined his diagnostic techniques become. The profiles described above (in the text) are a beginning in this direction. They are promising only to the extent that they are based upon a reliable and valid testing program which is far reaching in the outcomes it seeks to measure.¹

Following a suggestion of President L. D. Coffman the University Committee on Educational Research of the University of Minnesota appointed a Committee on Examinations in 1931. Since that time the practice of using new type tests to replace or to supplement the traditional examinations has been extended beyond the General College to other divisions of the University.

EXPERIMENTS WITH NEW TYPE TESTS: UNIVERSITY OF CHICAGO

The University of Chicago furnishes a third example of the extensive use of the new type test. Under a plan which was adopted in 1930 the University abandoned the peculiarly American organization of college education on the quantitative basis of credits. Students may now pass from the Junior to the Senior College as soon as they have shown that they have satisfactorily met the requirements prescribed for the Junior College entirely irrespective of the amount of time spent or the courses taken. They must pass examinations in four major fields: humanities, social sciences, physical sciences, and biological sciences. Class attendance is optional and examinations may be taken at any time when they are offered. Special importance is accordingly attached to the examinations, which are prepared with great care "to insure that the questions adequately represent the sort of mastery that the instructors have in mind and that the student is given ample opportunity to demonstrate his degree of mastery over a wide section of the course." The method of preparing the examination is similar to that used at the University of Minnesota except that the examiners themselves attend the lectures in the general courses and frequently consult the instructors, who in turn approve the questions in their subjects. The instructor may use tests in his course at any time, but the final comprehensive examinations are constructed

¹ A. C. Eurich, "Individual Diagnosis," in *Studies in College Examinations*, p. 86.

by full-time examiners. The system thus aims to divorce the function of examining, marking, and awarding credits from the function of instruction, a practice which in some way may be compared with the external examinations in England and France.¹ The forms of tests which are used are similar to those already mentioned; an attempt has been made to improve the essay form, but in this attempt as in all others with a similar aim the spirit of the essay disappears in favor of a closer approximation to the new type test.²

STATEWIDE TESTING PROGRAMS

One of the serious difficulties confronting American education is the absence of common standards of a comparable character throughout the country. There is no central authority for education either for the country as a whole or for the separate states which assumes the right to impose standards of attainment on educational institutions or to test for such standards. The tradition is strongly in favor of local autonomy, and, while such autonomy is salutary from one point of view—the adaptation of education to local environments—from another it has resulted in a great variety of standards which is only intensified by the fact that teachers, whose qualifications also vary throughout the country, conduct their own examinations in their own classes and that recommendations for admission to colleges are based on their results. With few exceptions, such as the College Entrance Examination Board and the New York State Board of Regents, the practice of external examinations for secondary school pupils is virtually unknown and is in any case suspect lest they control curricula, courses of study, and methods of instruction. There are, of course, regional standardizing agencies³ which have made important contributions to the definition of the scope of secondary school work, of its relations to colleges, and of standards of a quantitative nature, such as length of school year, length of class periods, size of staff, pupil load, preparation of teachers, teaching load, and buildings and equip-

¹ M. A. Desclos in summarizing the results of the French Committee's investigation on the scoring of the examination for the *baccalauréat* states as the first conclusion that "A selection should be made among examiners. The art of examining demands qualities which are quite different from those required for the art of teaching. One may be an excellent teacher without necessarily being a good examiner and there is no more reason for being offended on this account than because one is not a good musician." *La Correction des Épreuves Ecrites dans les Examens*, p. 379 (Paris, 1936).

² For a description and examples of the tests see the *Manual of Examination Methods* by M. W. Richardson, J. D. Russell, J. M. Stalnaker, and L. L. Thurstone (University of Chicago Bookstore, 1933).

³ Among these are the North Central Association, the Middle States Association of Colleges and Secondary Schools, the Association of Colleges and Secondary Schools of the Southern States, and the Northwest Association of Secondary and Higher Schools. See I. L. Kandel, *History of Secondary Education*, pp. 466 ff (Boston, 1930).

ment. The system of accrediting institutions and in some states inspection of high schools have contributed to the same ends. But in general it has been assumed that the quality of instruction could be depended upon if the quantitative standards were sound. There is further a surprising variety in the methods of admission to college and in the specific subjects required. The following table indicates the most common combinations of entrance criteria for regular students employed in 517 colleges:¹

NUMBER OF INSTITUTIONS INCLUDING CERTAIN ITEMS IN THEIR MOST COMMON COMBINATIONS OF ENTRANCE CRITERIA FOR REGULAR STUDENTS

<i>Criterion</i>	<i>Frequency</i>	<i>Rank</i>
Transcript of high school credits	318	1
Recommendation of principal	262	2
Personal interviews with applicants	138	3
Rank in high school graduating class	135	4
Recommendations by persons other than the principal	133	5
High school diploma	131	6
Character rating	126	7
Presentation of high school subject certificates	118	8
Examination devised and administered by institution	101	9
College Entrance Board examination	93	10
Intelligence test	93	11
College aptitude test	53	12
Other examination (Regents, State Board, etc.)	44	13

The subjects required for entrance also vary widely. "A few institutions have no subject requirements at all, a few others have requirements only in English, and others specify the subject matter for nearly all the 15 units which are usually required for admission."²

With such variation in the standards of the high schools and in the methods of admission the important issue has arisen of defining more adequately than has been done in the past the meaning of "fitness for college education." On the one hand, students have been admitted more or less automatically on graduation from accredited high schools, with the result that a large number, ranging from ten to thirty per cent, found themselves incapable of continuing their studies beyond the first year. On the other hand, the situation has been aggravated by the constant increase in the numbers of students proceeding to college after the War and particularly during the period of the depression. In private institutions selective barriers could be maintained; in publicly maintained institutions this has not been feasible partly for political reasons, partly because of the American tradition of equality of opportunity.

¹ United States Office of Education, National Survey of Secondary Education, *Bulletin*, 1932, No. 17, Monograph No. 10, *Articulation of High School and College*, p. 19 (Washington, D. C., 1933).

² *Ibid.*, p. 30.

To some extent the problem is being met in the high schools themselves, for example, in New York State and California, by the adoption of two types of graduation certificates, one attesting the completion of the course, the other entitling the student to admission to the state college. Another practice which is still in its initial stages is to establish within the college a system of counseling and advice which will enable the student to discover the types of courses from which he can profit.

There is greater promise, however, in another method of meeting the difficulties in this situation and that is the utilization of objective methods of discovering fitness for college studies—in other words, the application of methods of testing. Examinations of the new objective type are conducted in a number of states by organizations which are connected with the state university or with a number of higher educational institutions in a state; they are also administered by organizations which are national in scope. This development has the important advantage of establishing comparable standards over larger areas, of defining in a qualitative way educational standards for the high schools, and of meeting the objections to the traditional method of examinations. To meet the objection which is still raised against the coercive influence of external examinations it is argued that the new system does not prescribe a syllabus of studies to be followed by the schools and that the new type tests are based on the actual work done or professed to be done, and the courses of study and the textbooks used by the schools; that they furnish a more reliable basis for comparing schools because they are objective; and that they are fair to all students at the same stage of development because they have validity and are comprehensive samplings of the work accomplished.¹

Such tests have been employed not only to discover fitness for college education, but they have led to the discovery of large numbers of boys and girls who are intellectually able to continue their education but who for various reasons may decide not to do so. The tests can thus be used for the discovery of students of normal and superior ability at one end of the scale; at the other end they may be used to discourage pupils of low ability from proceeding beyond the high school.

Among the states which have developed systems of statewide testing of high school students, generally in their senior year, are Alabama, Colorado,

¹ There still remains the objection which comes from progressive educators that this system merely perpetuates existing definitions of secondary education and prevents radical experimentation with the secondary school curriculum. It is to be noted, however, that the thirty secondary schools which are engaged in such experiments are subjecting themselves to objective measurements in the work which they undertake.

Georgia, Indiana, Iowa, Kansas, Michigan, Minnesota, Montana, New Hampshire, Ohio, Pennsylvania, South Carolina, Texas, and Wisconsin.¹ Only a few typical examples can be discussed here.

STATEWIDE EXAMINATIONS. OHIO

The increasing demands for admission to college in Ohio led to the conviction that the problem must be considered not from the point of view of excluding students but from the point of view of developing measures for discovering those who are likely to profit from a college education. The problem in other words is to find the right kind of education for the students at the right time, and with the rapid social and economic changes the greatest need is a fundamental sorting out of the things most worth knowing about students. While high school marks over a four-year period would be predictive of success in college, such marks do not at present mean the same thing in all high schools. Hence as many different avenues as possible for securing reliable information about students must be explored through an extensive system of guidance, which would include: (1) An annual intelligence test of every high school student in the state in order to obtain a reliable measure of intellectual promise; (2) Annual transmutation to a common basis of the marks of all high school students in the last year of their course in order to yield a comparable prognostic value; (3) The assignment of a teacher in each school to serve as a guidance counsellor; (4) At least one guidance interview for each pupil once a year; and (5) The adoption of a uniform cumulative record blank with a time projection giving an account of a student's abilities, attitudes, plans, ambitions, etc. In the interests of each individual student and for the good of society the present selective and coercive measures must be surrendered in favor of adapting education to the needs of each student.² Since a college education for all is pedagogically unthinkable both because of the present organization of college education and because of the danger of vocational overcrowding, the following principle must be accepted.

In counting the cost of education, we should count the cost *to society*, not merely to the state, the college, the student, or his parents, for example, but to all of these. And, if the risk of academic mortality to an individual is great, we should either avoid the risk, eliminate it, or failing in that, hedge against the inevitable loss to the individual where he fails. As a

¹ See D. Segel, *National and State Cooperative High School Testing Programs*. U S Office of Education, *Bulletin*, 1933, No. 9 (Washington, 1933)

² See H. A. Toops, "A Minimal Guidance Program for Secondary Schools." *Ohio High School Bulletin*, No. 9 (Ohio State University).

corollary, the individual should accept his responsibility for his risk, but this is impossible unless he be informed of his risk.¹

There has accordingly been developed the practice of applying the Ohio College Association Intelligence Test to each student in the sophomore and senior years in high school.

The test consists of five basic tests, the time on each test being long enough to give a high reliability on each subtest, so that a pupil may be retested to determine whether he has a specific weakness in any of the subtest abilities—arithmetical reasoning, same-opposites vocabulary, grammatical analogies, number completion, and reading comprehension.²

The test, which has a validity coefficient of .55 to .60 in predicting freshman college success in liberal arts courses, is used by some fifty colleges and teacher training institutions in the State. The test has the merit not only of predicting college success and thereby discouraging from entrance students who are likely to fail, but also of encouraging students of ability who otherwise would not have continued their education. Ample statistical evidence is available to correlate ability as shown by the test scores with degree of success in college. At the same time it is recognized that the intelligence tests alone cannot tell the whole story about a student, since they fail "to measure important character traits—such as industry, persistence, conscientiousness, enthusiasm, seriousness of purpose, and those other factors which have a real bearing upon success." Hence the suggestion already mentioned of supplementing such test scores by information collected on a cumulative record card. It is, however, claimed that the results of research based on the tests are immediately translated into practice by the high school principals to whom they are circulated in advising students on their educational careers.

It might perhaps be objected that the same results could be achieved by the traditional examinations, but the objection ignores a number of facts. First of all the traditional examinations set up an arbitrary standard of "pass" or "fail"; the results do not furnish the basis for an exact comparison of a large body of students such as is provided when they are arranged on a percentile scale; nor do they indicate, as extensive intelligence tests over a large area may do in time, if not already, the kind of education from which a student is most likely to profit. Attention must again be drawn to the fact that no reputable expert in the field of measurement makes any deterministic

¹ *Ohio College Association Bulletin*, No. 61, p. 659.

² *A Suggested Program of Pre-College Guidance for High Schools*, p. 11 (Ohio State University Press, Columbus, O., 1929).

claims for intelligence¹ tests and that the need of as wide a range of information as can be secured, covering not merely scholastic ability but all those traits that are connoted by the term "personality," is coming to be widely recognized.

In addition to the college aptitude test conducted by the Ohio College Association the State Board of Education of Ohio conducts tests in the elementary and high schools which are used chiefly by schools outside of the large cities. The tests are given by the principals, superintendents, or teachers, who score them and return the tabulated results to the State Department of Education which reports the final results in such a way that each school can compare its own average score and the score of each pupil with the scores for the whole State. The tests thus serve the purpose of motivating and improving instruction and make possible diagnosis of difficulties and remedial treatment; the second of these purposes can obviously not be met by the general reports which are made on the basis of the traditional examination. The State Board of Education also conducts a general scholarship test for high school seniors, specially selected by their schools for scholastic ability, promise, and character. The results of this contest are transmitted in turn to the colleges and universities of the State.²

STATEWIDE EXAMINATIONS: IOWA

The Iowa Every-Pupil Achievement Testing Program was instituted by the College of Education and the Extension Division of the State University of Iowa in order to develop a regional system of coöperative testing for the high schools of the State and "to provide superior instruments for the measurement of educational achievement, to encourage better scholarship, and to accelerate improvement in the content and methods of high school instruction." While the existence of agencies for the production of standardized tests in various school subjects is recognized, it is felt that they are inadequate and scattered and sporadic in their use and not adapted to conditions within relatively restricted geographical regions. For the Iowa Every-Pupil Testing Program the following advantages are claimed:

It is an annual coöperative program conducted for schools operating under relatively homogeneous conditions in a restricted geographical area; all materials are expertly constructed on a uniform basis under central supervision, resulting in high consistency in quality and in form from sub-

¹ It is unfortunate that this term still continues in general use; the more recent term, "scholastic aptitude test," is more appropriate.

² D. Segel, *op. cit.*, pp. 26 f.

ject to subject; all materials are revised annually in order to keep them intimately related to the instructional programs in the schools and to keep them abreast of recent technical advances in test construction; norms are simultaneously established for all tests on the same group of pupils and schools, resulting in high comparability from subject to subject; and the very large numbers of pupils and schools involved make possible the establishment of specialized and localized norms far more reliable and meaningful than those provided for the usual standardized achievement test, including separate norms for pupils at different levels of general intelligence, and norms for a single composite measure established upon the results for all pupils in all subjects in each participating school.¹

The same tests are administered on the same day to all the pupils (more than 40,000) in the coöperating high² schools, and norms are established each year on the results from all the schools. The results of each year offer bases for comparison of each school with its previous records and with other schools and for following the progress of each pupil from year to year. The reports on the tests have the further value of directing attention to weaknesses in instruction; they are thus diagnostic in character and suggest the nature of remedial instruction needed. They may accordingly contribute to a program of educational guidance and improvement of instruction. For the research student they furnish material for the investigation of such pressing and crucial problems as "the relationship of class size and academic achievement, between size and type of school organization and results achieved, or between such factors as the training and experience of teachers and the achievement of their pupils." The experience of a few years has already led to some important conclusions on the content and methods of instruction.

Experience gained through this project during the last three years has yielded conclusive objective evidence that high school instruction is, in general, overly dependent upon the textbook, that mechanical, rote procedures tend to predominate in learning and in recitation, and that what pupils actually acquire as a result of instruction frequently consists, in large measure, of a mass of more or less meaningless verbalizations. Pupils tend to depend upon the recall of "pat" expressions taken from the textbook and upon words and phrases which have been memorized with no appreciation of their meaning or significance, rather than upon a reasoned understanding of the concepts involved. This type of learning has perhaps been encouraged in part through the use of poorly constructed examina-

¹ E. F. Lindquist, *The Iowa Every-Pupil Achievement Testing Program*, p. 6. *Bulletin of the State University of Iowa*, New Series No. 716, December 2, 1933.

² The testing program has now (1936) been extended downward to include the upper grades of the elementary schools.

tions of a highly informational character, such as have often been prepared by classroom teachers for informal testing purposes or such as have frequently been provided in the past by standardized test agencies.

It must be recognized that many standardized testing programs or other examination procedures—notably certain state board examinations—have tended to perpetuate these undesirable practices in teaching. The sponsors of the Every-Pupil Testing Program have been keenly aware of this truth and have been constantly on guard against dangers of this type. They believe they have demonstrated during the last three years that it is possible to conduct a regional testing program in such a way that these practices will be definitely *discouraged*, rather than perpetuated. The Every-Pupil tests are constructed with the conscious and deliberate purpose to *defeat* rote learning of, or mechanical and verbal drill upon, poorly selected and isolated information. In them, an effort is made to avoid the use of stereotyped textbook language; to call for *applications* rather than statements of laws and principles; to require the *interpretation* of real problem situations, rather than the repetition of words; to place the emphasis upon interpretative ideas rather than upon descriptive information only; and to require the pupil to draw inferences from, or to recognize the implications of, facts rather than merely to recall the facts themselves.¹

Sixty-minute tests are available in the following high school subjects:²

Ninth Year Algebra	World History
General Science	American History
First Year Latin	American Literature
English Correctness 9*	English Correctness 11*
Plane Geometry	Physics
Second Year Latin	Economics
Biology	American Government
English Correctness 10*	English Literature
English Correctness 12*	

**In this subject, the same test is administered in all four grades.*

Each school decides on the tests to be used but every pupil must be submitted in the subjects chosen. The tests are administered and scored in each school, to which a Summary Report of Results and a special confidential report is subsequently sent by the College of Education. The report shows (1) the average score made on each test in the school interpreted for each school in relation to rank position, T-score value, and size of school; (2) a composite measure of performance of the school in all subjects (rank posi-

¹ *Ibid.*, pp. 7 f. See also "A Cooperative Testing Program for Iowa Elementary Schools," pp. 12 ff. *Bulletin of the State University of Iowa*, New Series No. 760, October 6, 1934.

² *Ibid.*, p. 11.

tion, percentile position, etc.); (3) frequency distribution of average score for all schools, of composite measures, of the average scores of the schools in each enrollment group, and of the raw scores of individual pupils for all schools.

The significant characteristics of the norms furnished for the Every-Pupil tests are claimed to be as follows:

1. They are based upon a sufficiently large number of schools and pupils to establish their reliability beyond any question.
2. They are established only for schools within a restricted geographical area and therefore for schools that are operating under essentially the same conditions.
3. They are expressed in terms of school performance as well as in terms of pupil performance.
4. Separate norms are established for separate enrollment groups.
5. The norms are highly comparable from subject to subject.
6. Norms are established for composite school achievement as well as for achievement in individual subjects.
7. Separate norms are established for pupils at different levels of scholastic aptitude or general intelligence.¹

STATEWIDE EXAMINATIONS: WISCONSIN

Another example of statewide coöperation in conducting, as it were, a census of the intellectual resources in the youth of the state is furnished by Wisconsin, where it was found in 1930 that there had been an increase of over 400 per cent in the enrollment in high schools and colleges in twenty-five years as compared with an increase of only 33½ per cent in the population. This increase in the student body has been accompanied by "the serious tragedy of student mortality" with the consequence that

Everybody has considered whether it might not be possible to prevent some of the mortality by analyzing individual cases and trying to detect those who are doomed to failure if college work is attempted, by a discovery of those who have superior college ability and by analyses of individual

¹ *Ibid.*, pp. 16 f. The tests used in the Program are sold at four cents a copy, on the basis of three or four tests the cost for each pupil averages from twelve to sixteen cents, and a full testing program for a school of 100 pupils would be about \$15. No charge is made for descriptive pamphlets, manuals of directions, scoring keys, and the final summary reports. The cost of constructing and printing the tests, of preliminary experimentation with materials, of the statistical analysis of results, and of the preparation of the report is met out of income derived from the sale of tests. The cost of personnel is met by the University. The Program is, therefore, non-profit making. The College of Education also conducts the Every-Pupil Contest, a competitive test between schools; the State Scholarship Contest, a competitive test for individual students; and since 1934 the Iowa Every-Pupil Elementary School Testing Program, described in the *Bulletin of the State University of Iowa*, New Series No. 760, October 6, 1934.

interests and aptitudes which shall serve as a more intelligent basis for making life decisions.¹

The problems raised by this situation, affecting as they do the fundamental problem of the distribution of education "point the way to the need of experimentation" and constructive coöperation. To this end a State Committee on Coöperation was created in 1928 with representation from the High School Principals' Association, the Association of City Superintendents, the Privately Endowed Colleges, the State Teachers Colleges, and the State University, and with participation in an advisory capacity of the State Department of Public Instruction. The philosophy which guided the work of the Committee was defined as follows:

That educational opportunity shall be viewed as a broad highway extending from kindergarten to college graduation and that every child has a right to travel this highway as far as his native interest, capacity and endowment will permit. If this philosophy is sound then the committee felt that it must think of its problem, not in terms of that democratic principle which insists upon the political equality of human beings but in terms of the principle of biological *inequality* of human beings. While the committee subscribed completely to the principle of equality of opportunity it recognized just as completely that equality of opportunity means, not identity of opportunity but diversity of opportunity. The committee hoped that it might be successful in giving additional impetus to a program which should result in a keener realization by the schools of the state of the responsibility of education to analyze every boy and girl for native aptitude, and to know him as an individual, to the end that every student as he travels the broad highway of education may be assisted in an adjustment to that combination of circumstances in life which give, for him, the greatest promise of success and happiness.²

To carry out the program suggested by this philosophy, the Committee first secured an improvement in the admissions blanks used by colleges and universities in order that secondary school principals might be encouraged to assemble as much information as possible about each student and that college authorities might create a system of counselling and advice on the basis of this information. The improvement of the admissions blank led, secondly, to the appointment of a sub-committee to draw up a cumulative record card which "would tend to make judgments of pupils more objective." The third

¹ V. A. C. Henmon and F. O. Holt, "A Report on the Administration of Scholastic Aptitude Tests to 34,000 High School Seniors in Wisconsin in 1929 and 1930," p. 6. *Bulletin of the University of Wisconsin*, Serial No 1786 (Madison, Wisc., 1931).

² *Ibid.*, pp. 6 f

task was the coöperative testing program on the basis of which (1) comparable standards could be established for the high schools of the state, (2) ratings of ability to do college work could be stated in such forms as to enable a teacher to advise pupils and parents more intelligently, frankly, and confidently, and doubtful students could be advised against proceeding to college, while superior students who otherwise would not have continued their education could be encouraged to go on.

The Committee decided first to use the Ohio State University Psychological Test (see p. 106) because of its demonstrated reliability and validity (.55 to .60); from 1930 on the American Council Psychological Examination was used because its validity was as high as that of the Ohio test and because it was less costly. The cost of administration was divided between the participating colleges. The tests were conducted on one day, were administered by the high school principals, and were then sent to one of the colleges to be scored. The final check of the tests and tabulation of all data were made by the Bureau of Guidance at the University of Wisconsin. In 1929 the tests were taken by 16,619 seniors, out of an estimated total of about 17,000, in 434 public and 14 private and parochial secondary schools.¹ Reports on the results of the test were made to every high school and gave the percentile rank of each senior tested; similar reports were made to the colleges which the students had indicated as their probable choices.

The results of the tests applied to a supposedly homogeneous group—high school seniors—revealed an extraordinary range of individual differences:

While the ablest graduate makes a point score of 335 or over out of a possible 395, the poorest makes a point score of 30 or less. While the top five per cent make point scores of 240 or over, the lowest five per cent are making scores of 100 or less. Each of these are high school seniors, potential candidates for admission to higher institutions and further academic work.²

Obviously mere graduation from high school cannot be accepted as having any value for predicting success in college; it must be supplemented by some other information, an important part of which is furnished by a comparative test of the kind undertaken by the Committee in Wisconsin. Certainly the test used has sufficient reliability and validity to warrant a warning to certain groups of students on the percentile rank in the state of the likelihood

¹ In 1930 approximately 17,000 seniors were tested in 425 public and in 28 private and parochial schools; in 1931 a few over 20,000 seniors were tested in 433 public and 23 private and parochial secondary schools; and in 1935 the number of seniors tested in 428 public and 49 private secondary schools was 27,590; in addition 32,443 sophomores in 473 secondary schools, public and private, were also tested

² *Ibid.*, p. 20.

of failure should they undertake college studies. The test may also be used to encourage able students to pursue their studies further.

Out of 1,660 high school students in the top decile, there are 450 who express no intention to continue but for whom further academic training would be a good investment for society. The encouragement of such students in some way is clearly indicated. If we include the top two deciles, there would be 1,198 students who indicate no intention of going on to college but from whose further education society might profit a thousand-fold. The matter of selecting individuals for training is as important for society as proper training for them when selected. The encouragement of those who are best fitted for further scholastic training is more important than to discourage those who do not possess the capacities for success in college. It should be clearly indicated that no stigma attaches to such ratings as may be given by a test of scholastic aptitude. The test is not a test of general intelligence, whatever that may mean, but rather a test of that segment of general intelligence which is indicative of ability to profit by further scholastic training. The determination of individual differences as exactly as is possible at the present time and the assignment of individuals to the work for which they are best fitted is as important a problem as there is in the whole range of pure and applied science. To say that an individual is not possessed of the capacities necessary for successful college work is not to indicate that he may not be eminently successful in other fields of the world's work. The important thing is not to encourage an individual to engage in an occupation or set out upon an adventure in which he is probably doomed to failure.¹

In 1932-33 a test of mental ability began to be given to tenth year or sophomore students in the high schools. The statewide coöperative testing program in Wisconsin thus represents a twofold effort to secure comparable standards or ratings of ability of students at the end of their high school course and to use such ratings for the purpose of advising individual students in order that the number of failures in college may be reduced and the college student body in general improved.

STATEWIDE EXAMINATIONS · MINNESOTA

One other illustration of methods for meeting the large increase in the number of students flocking to the colleges is furnished by the University of Minnesota, which on the basis of the results of a testing program has under-

¹ *Ibid.*, pp. 21 f. It is unnecessary here to do more than to indicate other relationships which the coöperative test made it possible to study, such as "sex differences in scholastic aptitude among high school seniors", "occupational preferences of high school seniors in relation to scholastic aptitude", "intelligence and parental occupation," etc.

taken an experiment in education beyond the high school level for students who do not show promise of success in the courses regularly offered to undergraduates. The question "Who should go to College?" has been studied at the University of Minnesota since 1915, when a system of counselling was introduced in the College of Science, Literature and the Arts. Efforts were made to discover methods of predicting success in college, partly to save students from failing, partly to advise them in the choice of a vocation and course of training. College aptitude tests have been developed and used since 1917 and have involved the use of rank in high school and a psychological test. The average grades of each student during his high school career are calculated by the principals and then converted into percentile ranks to facilitate statewide comparisons. All students are required before entering college to take a psychological or college aptitude test¹ the score on which is

UNIVERSITY OF MINNESOTA				
COLLEGE SUCCESS MEASURED BY COLLEGE APTITUDE RATING				
<i>Freshmen Entering 1923 to 1927 Inclusive</i>				
(Number and Per Cent Doing Satisfactory Work)				
<i>C.A.R.</i>	<i>Number</i>	<i>Number Satisfactory</i>	<i>Per Cent Satisfactory</i>	
96-100	88	85	96.6	91.4
91-95	125	116	92.8	
86-90	156	121	76.3	71.9
81-85	135	98	69.5	
76-80	137	89	65.0	
71-75	123	69	56.1	42.0
66-70	131	57	43.5	
61-65	144	66	45.2	
56-60	144	57	39.6	
51-55	150	45	30.0	
46-50	141	34	24.1	18.0
41-45	129	20	15.5	
36-40	137	19	13.7	
31-35	116	13	11.2	9.2
26-30	90	6	6.4	
21-25	76	1	1.3	1.1
16-20	85	2	2.3	
11-15	47	...	...	
6-10	44	...	...	
1-5	14	...	...	
	2,212	898		

¹ This test was prepared for the Association of Minnesota Colleges, sixteen colleges including the University of Minnesota, which cooperate in administering and using the results of the test.

also converted into percentile rank. The combined rank on high school scholarship and on the test yield the College Aptitude Rating (C.A.R.).

Studies which have been continued since 1921 have shown that college scholarship could be predicted with a high degree of success from the C.A.R., which gives a better prediction of grades than either high school grades or the psychological test alone. The following table which gives the C.A.R. and the number doing satisfactory work is based on a study of 2,212 freshmen entering the liberal arts college of the University of Minnesota from 1923 to 1927 inclusive ¹

By satisfactory work is meant an average grade of C which is required for promotion to the senior college or for admission to a professional school. Of all the students who do satisfactory work 89.4 per cent come from the upper half of the college aptitude rating scale, whereas only 10.6 per cent come from the lower half. Of all the satisfactory students 99 per cent had ratings above 30. "Most of the students above 75 in the scale do satisfactory work while most of those in the lower part of the scale fall below the grade of C" In another report of a follow-up study of 1,085 freshmen of 1923, 1924, and 1925 it was found that only 6 of the 1,085 with a C.A.R. below 26 eventually graduated.² It is recognized, however, that valuable as the C.A.R. has been found to be in predicting college success, the information so obtained may well be supplemented with additional information on a student's interest, energy, industry, determination, application, and other qualities which in a very few cases appear to have enabled students to raise their college marks above the predicted level.

The logic of the situation led to the establishment in 1928 of what is known as the General College intended to provide a suitable education to students who according to the C.A.R. were not likely to succeed in the College of Science, Literature and the Arts. The General College represents an experiment in adapting education to the tested abilities of the students, or, in the words of President Lotus D. Coffman, it is

a new experiment, an adventure in the field of higher education. It is intended to provide a superior intellectual opportunity for a body of university students whose needs cannot now be adequately met by the existing organization of the University. It will succeed or fail in terms of its service to students. Its courses should be open to the most gifted student

¹ From a pamphlet *Who Should Go to College?*, p. 21, prepared by Dean J. B. Johnston (University of Minnesota, Minneapolis, 1930).

² J. B. Johnston, and E. G. Williamson, "A Follow-up Study of Early Scholastic Predictions in the University of Minnesota." *School and Society*, Vol. XXXX, December 1, 1934, pp. 730 ff.

in the University. Any student should be privileged to elect membership in the General College.¹

The General College offers a two-year course, leading to the degree of Associate in Arts. To obtain this degree students are required to pass a comprehensive examination in General and Contemporary Affairs Studies and five selected from the following: History and Government Studies, Economic Studies, Social Problems Studies, Psychology Studies, Euthenics Studies, Art Studies, Physical Science Studies, Biological Science Studies, and Oral and Written Communication Studies. By special permission students in General College may attempt courses in other colleges and departments of the University if they "wish to find out about their abilities in highly specialized fields of academic activity." They may also be permitted on the basis of their records to transfer to other colleges of the University and to enter senior colleges and proceed to a degree.

An investigation of the experiment from 1928 to 1932 showed that the students in General College who at that time were definitely of low aptitude rating obtained on the average higher grades than students with a corresponding rating in previous years who attempted the regular work of the regular college; their grades were higher in economics, history, and lecture courses in science than in languages or laboratory courses in science. Of 69 out of 169 students who in 1932-33 protested against assignment to the General College and were admitted to the Liberal Arts College on probation 28 remained for only one quarter, 7 for two quarters, 11 for three quarters, and 22 completed the second year; 10 obtained an average grade of C, and of those who did not obtain this grade 28 were dropped for low scholarship, 7 transferred to the General College, and one to the School of Nursing, 9 dropped out voluntarily, and 14 continued for the second year. In 1933-34 of 100 of this group admitted to the Liberal Arts College 1 cancelled during the first quarter, 47 remained for one quarter, 17 for two quarters, 35 for three quarters; of 48 who were dropped for low scholarship 39 entered the General College and 34

¹ Quoted in "General College of the University, 1935-36," p. 10, *The Bulletin of the University of Minnesota*, Vol. XXXVIII, September 4, 1935. The General College is open according to this *Bulletin* (pp. 9 f.) to (1) Those who desire to pursue courses or curricula not offered in other colleges. (2) Those who, for financial or other reasons, have only a limited time to give to college training. (3) Those who need and wish general orientation in the choice of, and general preparation for, a vocation. (4) Those who do not satisfactorily meet the entrance requirements of the other colleges because of lack of training in specific subjects. (5) Those who transfer from other institutions who do not meet the standards of advanced standing of the college to which they apply. (6) Those who are transferred by mutual agreement of the General College of the University and the college in which they propose to register or are registered. (7) Those who might not be accepted by existing colleges because of a lack of preparation to pursue their curricula.

failed to do satisfactory work as measured by a C grade. Of 53 students who transferred in 1932-34 from the General College to the Liberal Arts College 30 did satisfactory work and 23 did not.¹

While the General College is an interesting and, under the present socio-economic conditions which affect education, a desirable experiment in adjusting education to the needs and abilities of students, from the point of view of the present discussion it is mainly valuable as throwing light on the dependability of the system of measuring ability and predicting success in college. The general conclusion of the study conducted over a period of years is that

The C.A.R. offers a dependable and just basis for the counseling of students under conditions prevailing at such institutions as the University of Minnesota. Its use for the assignment of students to different colleges without counseling, however, is not recommended. If so used, it should be limited to those students who fall below the 15 or 20 percentile level.²

The essential question which arises out of the practice at the University of Minnesota and elsewhere is whether similar results could be obtained by the traditional examination method in which students are generally divided into the "pass" or "fail" groups, and in which the results, while in the main predictive for the best minds, furnish no clue to the educational abilities of the rest. It may further be objected that the action taken at the University of Minnesota in establishing the General College represents a dilution of standards of scholarship and of liberal education. The question raised by this objection is too important to be discussed here, but it may be countered with two other questions: First, whether exposure to a course in which such standards are maintained with the certainty of failure is true education for the students concerned, and, second, whether such standards cannot be better and more honestly maintained by providing for those who from the start are certain to fail such types of courses as are best suited to their abilities.

The fundamental question that is raised is whether under present conditions when students in larger numbers are entering and remaining in high schools and colleges the task of educational statesmanship is not the distribution of the right education for the right pupil under the right teacher rather than selection. For the University of Minnesota at any rate Dean J. B. Johnston gave the answer on the basis of about fifteen years of experience with the new system of classifying students according to their abilities:

¹ J. B. Johnston and E. G. Williamson, *loc cit*, pp 737 f

² *Ibid.*, *loc cit.*, p 738.

Three levels of ability are recognized in classifying students, and any individual who presents outstanding peculiarities is given individual treatment. The upper half of the class—that is, those entering freshmen whose college aptitude ratings are above 50—are admitted without any conditions or provisions, and they are permitted to register for any curriculum offered in the College of Science, Literature and the Arts. Individuals may be advised or required to take the course in “How to Study,” and all are classified in English on the basis of English training tests and their college aptitude ratings.

The second group includes roughly those whose college aptitude ratings are between 26 and 50. These students are classified in English and are advised to select studies that will have immediate value to them if they remain in college only a short time, and to postpone laboratory courses and languages.

What is only advice for the second group becomes positive direction for the third group.¹

NATIONWIDE TESTING PROGRAMS IN THE UNITED STATES

The absence of any official authority which as in European and other countries has the power to set up nationwide standards in education has already been mentioned. Nor, with one exception, does there exist, as in England, the system of external examiners through whom comparable standards may be maintained in the field of higher education. While there are certain advantages, such as flexibility, variety, and adaptation to local conditions, which may result from the absence of a system of standardization, with the large number of high schools and colleges and the unparalleled enrollments in these institutions the disadvantages are beginning to be recognized. Efforts have of course been made by the creation of accrediting agencies at both the secondary and college levels to secure some uniformity of practice, but in the main the standards which have been established affect externals—equipment, endowment, library and laboratory facilities, and size and qualifications of the faculty.² Equally mechanical and quantitative are the definitions of educational requirements in terms of units, points, and credits.

¹ J. B. Johnston, “Advising College Students,” *Journal of Higher Education*, Vol. I, June, 1930, pp. 317 f.

² The movement in the direction of setting up qualitative in place of quantitative standards is discussed in the *Manual of Accrediting Procedures* of the Commission on Institutions of Higher Education of the North Central Association of Colleges and Secondary Schools (Chicago, 1934). See also G. F. Zook, “Accreditation of Secondary Schools in the Light of the North Central Association Report,” pp. 76 ff. of the *Report of the Third Educational Conference* of the Committees on Personnel Methods and on Educational Testing of the American Council on Education, the Commission on the Relation of School and College of the Progressive Education Association, the Cooperative Test Service, and the Educational Records Bureau (Washington, D. C., 1934).

The results have been, first, that the standards of requirements and instruction are the standards of the faculty and of the individual instructors who conduct their own courses, and give and grade their own examinations in these courses, and, secondly, that there is no challenge to the student, once he has obtained the necessary credits in a course, to feel any responsibility for conserving whatever achievement he may have made in it. Attempts to correct the defects of these practices have been made in a relatively small number of institutions at the college level by the adoption of comprehensive examinations at the end of the course.¹ But in either case, whether graduation is based on the accumulation of a certain number of course credits or on the comprehensive examinations, there remains the assumption that students must follow and be examined in certain blanket prescriptions or mass requirements. Such a system, it is argued, sets up certain obstacles, arbitrarily defined in terms of general culture or liberal education or professional prerequisites, and tends to disregard the student as an individual.

COLLEGE SOPHOMORE TESTING PROGRAM

To meet the need of comparable standards at the college level and to promote a movement for greater individualization of education the Central Committee on Personnel Methods of the American Council on Education, a public but not an official body maintained by a number of colleges and universities throughout the country and by grants from some of the educational foundations, established an Advisory Committee on College Testing. Since it was early recognized by this Committee that any program of college testing must take into account the secondary schools from which college students are recruited, the name of the Committee was changed to the Committee on Educational Testing. In 1931 the Committee, influenced by the progress of the investigation being conducted by the Carnegie Foundation in Pennsylvania, recommended the adoption of a College Sophomore Testing Program, which was first used in May, 1932, in 140 institutions distributed in 38 states and including colleges of liberal arts, teachers colleges and teacher training departments, junior colleges, agricultural colleges, and engineering colleges. The number of students who took the test was 18,134, of whom 16,411 were sophomores.

According to the original proposal of the Committee the purpose of the College Sophomore Testing Program was defined in these terms:

The main purpose of the tests is to throw light on the capacities, needs

¹ See E. S. Jones, *Comprehensive Examinations* (New York, 1933).

and problems of individuals rather than to furnish a basis for institutional comparisons. The results of the testing program here suggested, when announced, may throw light on the selection of students and the conditions affecting the selection; on the response of students to the formal and informal facilities for cultivation offered by the college; on the relative effectiveness of instruction in departments; on the importance of various factors influencing student performance: age, home and social conditions, previous school training, scholastic and vocational aims, extra-curricular interests, faculty counsel and the like.¹

It was hoped that the use of a testing program through the coöperation of colleges distributed over a wide area would make available information that could be used for the successful guidance of individual students in order to provide for their better training for the service of society. In this program the Committee on Educational Testing coöperates with other agencies of the American Council, such as the Coöperative Test Service, and with the Educational Records Bureau, which will be discussed later. It was expected that as a minimal program the English, Intelligence, and General Culture sections of the Pennsylvania College Achievement Test, which the Carnegie Foundation placed at the disposal of the Committee, would be used. The following tests were available but not all were necessarily given in all the participating colleges:

1. Coöperative English Test, Series 1, Form 1933. A comprehensive English examination prepared by S. A. Leonard, V. A. C. Henmon, M. H. Willing, W. W. Cook, D. G. Paterson, and F. S. Beers, including sections on English usage (100 items, 65 minutes), spelling (55 items, 10 minutes), vocabulary (100 items, 20 minutes). Total, 255 items, 95 minutes.

2. Coöperative Literary Acquaintance Test, Form 1933. A literary acquaintance and literary information test prepared by F. S. Beers and D. G. Paterson (200 items, 45 minutes).

3. Coöperative General Culture Test, Form 1933. A cultural test, largely informational in character, prepared by E. F. Lindquist, H. R. Anderson, S. Timberg, J. Storck, and others, including three sections: history and social studies, 268 items, 80 minutes; foreign literature, 295 items, 55 minutes; fine arts, 240 items, 45 minutes. Total, 803 items, 180 minutes.

4. Coöperative General Science Test for College Classes, Form 1933. A brief general science test prepared by R. W. Tyler, H. W. Farwell, G. M. Kay, W. J. Eckert, S. R. Powers, F. L. Fitzpatrick, O. E. Underhill, V. H. Noll, and A. W. Schindler; 270 items, 60 minutes.

5. Coöperative General Mathematics Test for College Classes, Form

¹ *Educational Record*, Vol. 12, July, 1931, p. 346.

1933. A comprehensive test of high school mathematics, with brief sections on college algebra, analytics, and calculus; prepared by H. T. Lundholm and L. P. Siceloff; 166 items, 120 minutes.

6. A Foreign Language Test: French, German, or Spanish (reading 80 items, vocabulary 100 items, grammar 100 items) or Latin (vocabulary 100 items, grammar 100 items, reading 137 items); total time for each language test, 90 minutes.

7. Coöperative Contemporary Affairs Test, by Alvin C. Eurich, Elmo C. Wilson, and others. This test is based on an analysis of the content of current literature and is intended to measure the extent to which students keep abreast of significant current affairs in the fields of art, literature, government, international relations, politics, economics, religion, etc.; 207 items, 90 minutes.

8. Coöperative Physics Tests, by the Committee on Tests of the American Association of Physics Teachers. Part I, Mechanics, 60 minutes; Part II, Heat, 30 minutes; Part III, Sound and Wave Motion, 20 minutes; Part IV, Light, 40 minutes; Part V, Electricity, 50 minutes; Part VI, Modern Physics, 25 minutes. The six units together constitute a general comprehensive test in college physics, but each college was advised to choose a combination of topics to suit local needs.

9. Cooperative Chemistry Tests, by B. C. Hendricks, F. P. Frutchey, A. G. Horney, R. W. Tyler, and others. Part I, Information and Terminology, 90 minutes; Part II, Interpretation of Experiments, 45 minutes; Part III, Application of Principles, 45 minutes; Part IV, Symbols, Formulae, and Equations, 75 minutes; C., a combined test covering three important aspects of elementary chemistry, Information, Terminology, and Application of Principles, 2 hours.

10. Coöperative Zoölogy Tests, by W. M. Barrows, J. W. Price, L. H. Snyder, and others. Part I, Information and Terminology, 90 minutes; Part II, Interpretation of Experiments, 45 minutes; Part III, Application of Principles, 45 minutes; Part IV, Identification of Structures and Their Functions, 75 minutes; Part V, Laboratory Techniques: (1) Using the Microscope, (2) Making a Temporary Mount of Frog's Skin, (3) Making a Fresh Mount of Striated Muscle, (4) Dissecting Muscles of a Frog's Hind Leg, (5) Dissecting the Ventral Body Wall of a Frog, (6) Dissecting the Digestive Tract of a Frog, (7) Dissecting the Circulatory System of a Frog, (8) Dissecting the Brain of a Frog, (9) Dissecting the Earthworm; C., a combined test covering three important aspects of elementary zoology, Information, Terminology, and Application of Principles, 2 hours.

11. Coöperative Botany Tests, by F. K. Butters, P. O. Johnson, R. B. Gordon, C. W. Horton, H. C. Sampson, L. H. Tiffany, and others. Part I, Information and Terminology, 90 minutes; Part II, Interpretation of Ex-

periments, 45 minutes; Part III, Application of Principles, 45 minutes; Part IV, Identification of Structures and Their Functions, 75 minutes; Part V, Laboratory Techniques, Using the Microscope; C., a combined test covering three important aspects of elementary botany, Information, Terminology, and Application of Principles, 2 hours.

12. Cooperative Geology Tests, by G. Marshall Kay and others. Historical Geology, 90 minutes; Physical Geology, 90 minutes.

The detailed analysis of the results of the tests for each college is left to those most familiar with their own local situation, since the value of the results rests on the way in which they are used by the individual institution in guiding each student effectively. On the basis of a tabulation of all the results on national sophomore percentile scales the national standing of each student can be determined. The most important result which has been revealed since the testing program has been used is the existence of an extraordinary degree of variability among the colleges themselves and among the students in each college on every test taken. Not only is there variability among the students of one year but there has also been discovered considerable overlapping among all the students of all the four years, and of students in independent or private secondary schools to whom similar tests have been given by the Educational Records Bureau. The variability and overlapping can only signify one thing, that the attempt to deal with all students at any level as a homogeneous group has been proved to be unwarranted. The results of the tests have made it possible to study the prevalence and degree of variability in a large number of items—on each of the tests of intelligence and achievement—and to compare students studying for different degrees, general academic and professional.

In general the results of the tests conducted since 1932 tend to shake the traditional interpretation of equality of opportunity as identity of opportunity.

It is obvious that with such student bodies as here indicated, mass standards and "average performance" are not only irrelevant but ironical. It seems equally obvious that no simple and easy changes can be remotely adequate for making the readjustment demanded by the heterogeneity here exposed. Nothing short of a definite abandonment and repudiation of blanket prescriptions and unmitigated mass standards, and a genuine and wholehearted acceptance of the fact of individual differences and their increasingly obvious administrative and guidance implications, would seem to be a sufficient reason for hopefulness in regard to this crucial social problem.¹

¹ "The 1934 College Sophomore Testing Program," *Educational Record*, Vol. 15, October, 1934, p. 492.

The ideal of equality of opportunity has in education been responsible for pressure toward conformity, the development of uniform patterns more or less equalized and made identical on a quantitative basis, and insistence on common standards. The results of the tests disprove completely the possibility of attaining such an ideal and, if anything, point to the danger of producing a dead level of mediocrity. But "mediocrity, as an ideal, is not democratic. Mass instruction designed to level individual differences cannot show the schools what to do for the unusual or failing student."¹ It is obvious that this conclusion does not and cannot mean the adoption of a dual system of education—one for the able and another for the dull. The significance of such studies as the College Sophomore Testing Program would be lost unless it were realized that the fundamental conclusion is that education must be individualized and adapted to the abilities of each student.

Democratic ideals, far from aiming at a dead level of mediocrity, recognize and accept individual differences and would make use of them for socially acceptable ends. True education makes distinctions among men; it means finding one's own level, but misguided theorists on democracy and democratic education often obscure this fundamental truth. It is the function of the Committee (on Educational Testing) to help make the philosophy of individual guidance practical and in so doing to support the widespread movement toward revaluing educational traditions.²

The Committee is at pains in each of its reports on the College Sophomore Testing Program to stress a number of points. First, the tests may in time be improved from the point of view of validity. Second, tests can be improved only through the coöperation of experts in measurement and psychology and expert teachers in the subjects to be tested. Third, the tests do not and cannot test the whole of education and of instruction. Fourth, the tests furnish a certain amount of information which is valuable for the purpose of guiding individual students, but for effective guidance such information must be continuous and cumulative. Fifth, systematic records must be kept over a period of years and should include information not only of the kind that can be secured by means of tests, but of a far wider range so as to cover a student's peculiar character traits, his degree of initiative, his ability to carry out independent, self-initiated study, his health and other personal circumstances that may have a bearing on his work.

Reports from the colleges which have participated in the testing program indicate that many have already begun to use such systematic cumulative

¹ *Ibid.*, p. 477.

² *Ibid.*, p. 471.

records, some have embarked on programs of guidance and counseling, and others are beginning to make special provisions for students of different abilities. The ultimate aim of the testing program is to put into the hands of teachers and administrators of educational institutions information which will enable them to provide "the right education for the right pupil" or to quote from one of the Reports on the College Sophomore Testing Program:

Teaching, like the practice of medicine, is an art. The physician makes use of chemical analyses, fluoroscope examinations, and the like; the teacher uses achievement tests, personality ratings, and narrative accounts of school and work history. In the final analysis both teacher and physician prescribe or advise intelligently only in terms of the total clinical picture—not in terms of any single test or interview.¹

COÖPERATIVE TEST SERVICE

The development and progress of the College Sophomore Testing Program were made possible by the creation under the American Council on Education of the Coöperative Test Service. This Service arose out of the proposal of a Committee on Vocational Guidance appointed in 1923 by the National Research Council to consider methods for securing information about individual students in order to discover their fitness for different vocations. This committee soon became the Central Committee on Personnel Procedures of the American Council on Education. After a number of conferences four sub-committees were appointed in 1927 as follows: Sub-Committee on Personal Record Cards, Sub-Committee on Achievement Tests, Sub-Committee on Personality Measurement, and Sub-Committee on Occupational Information, to which another Sub-Committee on Personal Development was added about a year later. The work of these sub-committees was placed under the supervision of a Central Committee of the Committee on Personnel Methods of the American Council on Education. In 1930 it was decided that a definite program for the preparation of achievement tests could be prepared and developed. To this end a grant of \$500,000 was obtained from the General Education Board for the creation of a bureau, the Coöperative Test Service, to prepare over a period of ten years a series of tests adequate in number, sound in construction, and of measurable validity and comparability.

The essential feature of the Test Service is that it is coöperative. It enlists the coöperation of experts in universities throughout the country; it coöperates with schools and colleges throughout the country in administering the

¹ *Ibid.*, pp 476 f.

tests; it is subject to constant and continuous criticism which¹ makes the improvement of the tests possible. Standardized tests have been available for many years and in many but not all subjects in which objective examinations are desirable and needed, but these tests have been prepared by one or two individuals without adequate facilities for experimentation or critical evaluation. The Coöperative Test Service was established as a non-profit organization to produce a series of valid tests for a number of subjects of the high school and college curriculum in ten or twelve equivalent forms. In the words of the Director of the Coöperative Test Service, Professor Ben D. Wood, of Columbia University:

The main purpose of this Coöperative Test Service is to supply the colleges and schools of the country with measuring instruments which will yield accurate and comparable measurements of defined achievements.¹

The desirability of having equivalent forms of tests in the same subject was emphasized in order to prevent coaching in the schools and the use of tests as teaching devices. The fear has, indeed, been expressed that the extension of the system of testing over a large area would exercise a constricting influence on the work of schools and colleges. Experience has not justified these fears, in the first place because "the tests in most of the subject matter fields are based largely upon certain fundamental common aspects of the subject matters which are widely recognized as important elements of achievement in those fields, and therefore worthy of being measured."² One point that is reiterated by those who are most informed on the subject is that tests do not and cannot measure all the desirable aims of education, the more imponderable aspects of achievement, for which subjective judgments must continue to be relied upon.

All recognize the higher creative values of learning in English, foreign languages, and mathematics, and all recognize that the existing tests do not measure these values directly; but most schoolmen recognize that the aspects of achievement measured by the tests are positively correlated with the higher values, and an increasing number believe that their estimates of and control over the higher values are enhanced by the availability of comparable measures of the more objective aspects of achievement.³

Secondly, the fear of the constrictive influence of tests is not justified if it

¹ "The Cooperative Test Service," *Educational Record*, Vol. 12, July, 1931, p. 215.

² H. F. Hawkes, "Report on the Coöperative Test Service," *Educational Record*, Vol. 15, July, 1934, pp. 363 f.

³ *Ibid.*, p. 364.

is recognized that their purpose is to furnish sources of information regarding the achievement and growth of individual pupils. They do not, like the traditional examinations, merely separate those who pass from those who fail; they seek also to discover both the range of abilities and deficiencies in order that both may be improved. As stated frequently by Dean Max McConn they constitute for the educator a scientific measuring tool comparable to the thermometer and stethoscope of the doctor.

The scope of the work which has been set up for the Coöperative Test Service is indicated by the following analysis of the minimum steps required for the production of five equivalent forms of a 200-question test:

1. Survey field to be covered.
2. Construct 2,000 questions.
3. Edit these, eliminating about 400 questions by examination.
4. Print eight 200-question lists as nearly balanced in subject matter and as nearly equivalent in difficulty as possible by inspection.
5. Administer each of these lists to about 500 to 800 representative students, giving them in pairs to the same students; for example, AB, AC, AD, AE, AF, AG, AH.
6. Score these 4,000 or more preliminary question lists.
7. Analyze the difficulty and validity of each of the 1,600 questions.
8. Choose the 1,000 best questions.
9. Divide the 1,000 best questions into five balanced and equivalent tests, Forms A, B, C, D, and E.
10. Administer again in pairs to insure equivalence, to test reliability and validity, to correct and complete keys, and to secure adequate norms.¹

No better statement could be made than is implied by this analysis of the fact that examining is a major function apart from instruction. While the drafting of some ten questions in the traditional examination does not involve a great deal of time nor the marking require long conferences, experimentation and discussion, the major part of the time in the objective tests is taken up by the preparation and construction while the scoring is relatively mechanical.

The advantages claimed for the series of comparable tests, such as those prepared by the Coöperative Test Service, are that (1) they make possible the measurement of individual growth year after year in certain types of achievement; (2) on the basis of the results educational guidance and research may be conducted since the tests exist in a large number of comparable forms related to each other; (3) an educational institution can maintain

¹ Max McConn, "The Coöperative Test Service," *Journal of Higher Education*, Vol. II, May, 1931, p. 230.

standards of admission, placement, promotion, certification, and graduation which remain uniform from year to year; (4) by using the results of the tests as a common denominator the results of subjective and other local examinations can be made closely comparable.¹ The Coöperative Test Service publishes norms for its tests as they become available, subject to revision from time to time as additional data are secured. Tests for high schools and colleges have now been constructed, each in a number of equivalent forms, in English, foreign languages, mathematics, general science, botany, chemistry, physics, mechanics, heat, light, electricity, modern physics, geology, historical geology, physical geology, American history, medieval history, modern European history, English history, world history, economics, contemporary affairs, and general culture.²

COLLEGE PSYCHOLOGICAL EXAMINATION

Psychological tests for college freshmen, first introduced under the auspices of the National Research Council and developed in 1924 with the financial support of the Commonwealth Fund, have been conducted for a number of years under the American Council on Education and after two years became self-supporting. The conduct of these tests by Professors L. L. Thurstone and Thelma G. Thurstone of the University of Chicago has in itself been a test of the tests, aimed to secure those which are best from the point of view of diagnosis and prediction. The purpose of this annual test, now known as the American Council on Education Psychological Examination, is to secure not a single index of mentality but a profile chart which will show the special talents and deficiencies of each individual student and also make possible the derivation of a single composite index. On the basis of the results reports are prepared which give tables showing the norms of performance in each test, frequency of distribution, and gross scores. The test scores are correlated with academic grades of the students. In 1935 the American Council on Education Psychological Examination was given in 493 institutions to 189,506 students, although the tabulations were made only of the records of 58,402 students in 266 colleges. The following types of tests were used: completion, arithmetic, artificial language, analogies, and opposites.³

¹ The Coöperative Test Service, *Announcement of Tests*, November, 1935.

² For descriptions of the methods of constructing tests in the social studies and the considerations that must be taken into account see E. F. Lindquist, "The Form of the American History Examination of the Coöperative Test Service," *Educational Record*, Vol. 12, October, 1931, pp. 459 ff., and E. F. Lindquist, and H. R. Anderson, "Achievement Tests in Social Studies," *Educational Record*, Vol. 14, April, 1933, pp. 3 ff.

³ A Report on the Psychological Examination of each year is published in *The Educational Record*. The

EDUCATIONAL RECORDS BUREAU

Private or independent schools in the United States have on the whole enjoyed greater freedom than in other countries. They are not governed as in the countries of Europe and elsewhere by official regulations with reference to establishment, qualifications of teachers, and standards of instruction, nor have they subjected themselves voluntarily as in England to the requirements and standards of external examining bodies except at one point. The exception has been the necessity of preparing their pupils for admission to colleges, since the majority of the independent schools are predominantly college preparatory schools and the majority of their pupils do in practice plan to proceed to college. Hence, in so far as there is a common test it has consisted of the examinations either of the College Entrance Examination Board or of those colleges which have their own requirements for admission. During the last decade it has begun to be realized that these examinations, held at the close of a pupil's secondary school career, come too late to make constructive educational guidance possible. This realization was accompanied by a desire for reliable and comparable measurements of the native capacities and scholastic achievements which with other available information could serve as a basis for continuous study and educational guidance of pupils. Admission to college and the ultimate choice of a career were considered to be issues too critical to be subjected to the hazard of a single examination. In 1927 a number of independent schools decided to organize the Educational Records Bureau to assist member schools to secure reliable and comparable measurements of the native capacities and academic achievements of their pupils as a basis for continuing study over a period of years.

The Educational Records Bureau was incorporated and chartered under the Board of Regents of the University of the State of New York as a non-profit making organization, with its headquarters in New York. The Bureau is supported by an annual membership fee from the schools of \$15 and a charge based on service costs for each test supplied; the service charge may be reduced if member schools do not avail themselves of the full services offered by the Bureau. These services include scoring and reporting the results of the tests; comparative reports giving the distribution of pupil scores and percentile norms on the tests for the following classifications: (a) nationwide, (b) state and regional, (c) community and school according to size and

latest, "The 1935 Psychological Examination," appeared in *The Educational Record*, Vol. 18, April, 1936, pp. 296 ff. In 1933 a supplementary test was prepared for grades 9 to 12 but has not achieved the popularity of the original test, probably because of the existence of other local systems for conducting testing programs.

type, (d) sex, and (e) age; and confidential reports indicating the rank of the participating schools in the classifications mentioned ¹

It was planned at first to give tests only at the end of the junior year of the secondary school or the eleventh year of a pupil's scholastic career. After one year, when it was realized that still more information was needed than could be obtained from one series of tests, it was decided to institute examinations at the end of the tenth year. Finally, when the real purposes which such tests could serve were recognized, many of the schools adopted the practice of cooperating with the Educational Records Bureau in conducting tests for each pupil in each year of his career in school. The Bureau does not itself construct tests but employs tests prepared and supplied by the Cooperative Test Service, other organizations, and individuals. The Bureau's functions are chiefly to score and report on the results of the tests and to coordinate these results on a comparable basis. Some of the member schools undertake to score the tests themselves, but this practice defeats one purpose of the organization which is the establishment of significant norms representative of the whole group of schools, which now number 261.

Although the derivation of representative norms is desirable, it has never been intended that the tests should be regarded as a sort of interscholastic competition. Nor in practice have the tests been used in this way, although comparisons with the norms may undoubtedly lead to serious investigation of such weaknesses as may be revealed in a school,—the curriculum, the methods of instruction, or the classification of pupils. The paramount purpose of the testing program of the Educational Records Bureau has been to supply reliable objective and comparable data on the basis of which the pupils may be studied as individuals and adjustments of school activities be made in accordance with their needs. The results of the tests are used for corrective and remedial work; they may be used to discover whether pupils are working in accordance with their abilities; they serve as measures both of achievement and of educability. The tests differ from the traditional examinations in a number of respects: they are comparable from year to year; the results lend themselves to more detailed and realistic analysis of abilities and deficiencies; they can accordingly be used to discover the actual needs of each pupil as an individual. While it may be claimed that external examining boards have always published reports on the strong and weak points revealed in the examinations in different subjects, such reports have been collective in character; they have been directed at all the candidates examined but

¹ See D. Segel, *National and State Cooperative High School Testing Programs*, p. 5 U. S. Office of Education, *Bulletin*, 1933, No. 9 (Washington, D. C., 1933)

without revealing any particular details about each one; at best all that could be known about a candidate would be that he had passed or failed and the mark obtained, but so far as guidance for further education is concerned little or no help is supplied.¹ For the tests it is claimed that the results furnish a better understanding of each pupil as an individual and make possible closer adaptation of school procedure to his needs.

To this end the Educational Records Bureau has enabled schools to supplement the original testing program, normally conducted at the end of the school year, with another series of tests, given at the beginning of the school year, to test scholastic aptitude or intelligence and for purposes of diagnosis and placement. The scholastic aptitude tests which have been used are the American Council Psychological Examination, Grades 9-12; the Pintner-Cunningham Primary Mental Test, Grade 1; the Otis Group Primary Intelligence Scale, Grades 2 and 3; the Otis Self-Administering Intermediate Examination, Grades 4-7; and the Otis Self-Administering Higher Examination, Grade 8. The diagnostic tests include the Nelson-Denny Reading Test, Grades 9-12; the Iowa Silent Reading Test, Elementary Examination, Grades 5-8; and the Progressive Achievement Tests, Grades 1-12.

It is not claimed, however, either that the tests have reached ultimate perfection or that their results can tell the whole story about an individual pupil. One result of the testing program of the Educational Records Bureau has been to direct attention to the necessity of maintaining continuous records for each individual which should contain not only information about his ability as measured by intelligence tests, and achievements as measured by scholastic tests, but also records of teachers' judgments and marks, and of his physical and personal development and social adjustment. Inevitably it has come to be recognized not merely that tests constitute a better examining device but that they play an important part in a program of educational guidance. In other words the educational progress of an individual is not dependent merely on his intellectual or scholastic development but involves all those factors which go to make up the whole individual. Unlike the traditional examination which reveals in the main what a pupil cannot do, the testing program seeks to discover what he can and should be encouraged to do, *but this end can be achieved only by such continuous study of the pupil as is becoming increasingly possible by the use of cumulative record blanks.*

A long-standing grievance of college preparatory schools in the past has

¹ The same remarks would apply to reports sent to parents in such terms as "Could do better, if he tried," "Does not work up to his ability," "Does not concentrate sufficiently," or even "Is incorrigibly lazy." See also p. 143.

been that the external examinations have tended to dominate the curriculum and instruction.¹ The same fear has been expressed with reference to a testing program, that the tests would in the long run tend to define or "freeze" the curriculum and prevent flexibility and adaptations to changing demands. Although the program of the Educational Records Bureau has been under way for less than ten years, there is sufficient evidence to prove that this fear is unwarranted. Far from limiting the curriculum or mechanizing the educative process, the tests have served as instruments for a constant and critical evaluation of the curriculum; they have led to the elimination of parts that were unattainable and unprofitable; they have made its administration more flexible so far as the needs and abilities of each pupil are concerned. More than this, however, those most conversant with the purposes of the testing program have been the most assiduous in their insistence that the tests are not the whole of education, a point of view which may be contrasted with the place of the traditional examinations which too frequently occupied the center of attention. In other words, no more is claimed than can actually be contributed by the tests to a larger program of educational guidance.²

THE PENNSYLVANIA STUDY

The value of the new type objective examination has nowhere been more successfully illustrated than in the use to which it has been put in the Study of the Relations of Secondary and Higher Education in Pennsylvania. The widespread extension of the use of the new type examination in order to se-

¹ As a result of protest against the control exercised over the work of secondary schools by college entrance requirements a small group of public and private secondary schools, under the Commission on the Relation of School and College of the Progressive Education Association, was set free in 1932 to engage in experimental work on the understanding that a number of colleges will admit students from these schools for a period of five years, beginning in 1936, without further examination. It is interesting to note, however, that these schools, thirty in number, have recognized the necessity of devising some methods for appraising and evaluating their own work and the work of their pupils and will, among other items of information, submit to the colleges students' scores on scholastic aptitude, achievement, and other diagnostic tests given by the schools during the secondary school course, specialists in tests and measurements have been appointed to construct tests in accordance with the objectives and content of the experimental work which has been undertaken. See W. A. Aikin, "The Purposes of the Eight Year Experimental Study," *Educational Record*, January, 1935, pp. 107 ff.; and *Report of the Third Annual Conference of the Eight Year Study*, 1935.

² It may be of interest to note the wider scope of a testing program such as that conducted by the Educational Records Bureau. In 1929 a comparative study of English schools was made by the Bureau using intelligence tests and achievement tests in English, French, and algebra. See *Educational Records Bulletin*, No. 5, April, 1930. *Testing School Achievement in England and America. Report of the Results of Intelligence and Achievement Tests given in Four English Schools Compared with Records on Similar Tests of Students in American Independent Secondary Schools*, May, 1929. Reference may be made to other international comparative studies of achievement. S. R. Powers, "Chemistry in English and American Schools," *School Science Review*, Vol. VIII, June, 1927, pp. 253 ff., and a "Comparison

cure continuous information about pupils and students has been indicated in the preceding sections, but, while the need of such information for purposes of educational guidance has been recognized and admitted generally, its accumulation has rarely been continuous and in the hands of a single authority. The Pennsylvania Study derives its significance from the fact that continuous records of ability and achievements have been secured and maintained by the same body for a large group of students from their entrance into high school until their graduation from college, that is, over a period of eight years.

This is not, however, the only significant aspect of the Study, nor, important as it is, is the proof of the great degree of variability among students and the educational institutions which they attend and which undertake to treat them all very much in the same way. Apart from the major contribution of the Study as an illustration of the ways in which the new type examinations may be used, the importance of the Study rightly interpreted will lie in the attack which it has made on the characteristically American organization of education in terms of time spent in an educational institution or on a course, that is, the system of measuring education in terms of units, points or credits. Stated in another way, the Study has revealed the fact that the system of credits militates against the intellectual growth of students and the accumulation of enduring knowledge.

The Study arose out of discussions of the relations between high schools and colleges which had been conducted in 1925 and 1926 by the Association of College Presidents of Pennsylvania. A Commission, representing the Association and the State Department of Public Instruction of Pennsylvania, invited the Carnegie Foundation for the Advancement of Teaching to undertake an investigation of the problems involved in the relations of high schools and colleges. Interest in these problems had already been manifested in a study prepared by Dr. William S. Learned, staff member of the Carnegie Foundation, and published under the title of *The Quality of the Educational Process in the United States and Europe*.¹ The main thesis of this report was that the difference in the quality of education in England, Germany, and France as contrasted with the United States is due to the recognition of distinct aims, continuity of courses, and responsibility on the part of the students

of Content and Accomplishment of Students in English and American Secondary Schools," *Journal of Chemical Education*, Vol. IV, December, 1927, pp. 1505 ff., G. MacGregor, *Achievement Tests in the Primary School A Comparative Study with American Tests in Five* (London, 1934). Similar comparative studies have been made in British Columbia and Australia. The question may well be raised whether such comparisons could have been instituted with any degree of reliability if the traditional form of examinations had been used.

¹ Carnegie Foundation for the Advancement of Teaching, *Bulletin Number Twenty* (New York, 1927). This important Bulletin is unfortunately now out of print.

for the accumulation of acquired knowledge. While secondary and higher education in the European countries is dominated by a systematic effort to secure and promote continuity of intellectual growth along a number of distinctive lines, the program of students in corresponding institutions in the United States is made up of a number of unrelated, discrete parcels of knowledge or courses, each one on its completion receiving the sanction of a certain number of credits, which when piled up to the required amount lead to graduation whether from the secondary school or the college. At no stage, unless it be in a college entrance examination, is the student held to an accounting of the whole sweep of education that he has received up to that point.

The Carnegie Foundation agreed, with the coöperation of the colleges, the public high schools, and the State Department of Public Instruction of Pennsylvania, to undertake a study of the relations of the secondary schools and colleges over a period of seven years. Dr. William S. Larned was in charge of the investigation and associated with himself Professor Ben D. Wood of Columbia University as the expert in the technical field of tests and measurements. The plan of the inquiry was to subject the educational procedure of secondary schools and colleges as a unit to examination at three points: (1) at the time of entrance to the secondary school, (2) at the close of the secondary school course, and (3) toward the close of the college course. The Study, in other words, was intended to secure as much information about the scholastic and other progress of students from the last years of the elementary schools, through the secondary schools, and in the colleges up to the time of graduation. It would thus follow the progress of students through their educational careers and close with an assessment of the final product or "an inventory of the baccalaureate mind"¹ In this way it was hoped to discover as much information as possible about a whole school and college generation, and to answer some of the following questions: What sort of mind goes to colleges, teachers colleges, business, industry, etc.? How many students with marked aptitude for intellectual pursuits are unable to go to college for economic or other reasons, and what proportion actually enters college without evident qualifications? What sort of mind goes to college, continues, and grows in knowledge and power?

Two important instruments have been used in the conduct of the Pennsylvania Study, intelligence and new type tests and the cumulative record card on the basis of which students could be placed more accurately and ad-

¹ This plan may be compared with the prospective and retrospective studies undertaken by the Scottish Council for Research in Education and published in two volumes (1) *The Intelligence of Scottish Children A National Survey of an Age-Group* (London, 1933); and (2) *The Prognostic Value of University Entrance Examinations in Scotland* (London, 1936).

vised to the best advantage. Thus several thousands of pupils in the seventh year of their educational careers were tested for mental, social, and physical traits in order to obtain and record a comprehensive and trustworthy story of their achievement. The subsequent career of each of these pupils was followed on a cumulative record card which aimed to bring together information desired and to test experimentally the value of such record in all intellectual education. Through achievement testing and personal scrutiny the intellectual characteristics of the group of pupils concerned were unusually well understood. Three results followed from this part of the Study. The first was an emphasis on the importance of competent guidance of pupils on the basis of all of the information obtained about them. The second was the revelation of the inadequacy of a system of education based on credits and points instead of the provision of opportunities for a continuous program of education in accordance with the abilities and interests of a pupil. The third result was the corollary of the second, that the educational program of each pupil should be sequential and continuous instead of a collection of separate and discrete units, and should emphasize cumulative achievement to be measured, whenever the need might arise, by objective tests of a comprehensive character.

From the results obtained information was made available on the persistence of certain abilities, the development of new interests, and the effect on ultimate achievement of different types of curricula and school organization. Such information, it was found, would have the further effect of simplifying the relations of secondary schools and colleges, since the colleges would receive more continuous and comprehensive information about candidates for admission than has been available in the school certificates or than could be secured by a single entrance examination. Thus it was found in 1932 in following up the records of students who had entered college only to withdraw during the first year that the causes of what were apparently cases of maladjustment were the results of discrepancies between marks obtained in high school and scores obtained in objective tests, and ignorance of the ability, defined achievements, and effective interests of the individuals concerned.

When the investigation was carried into the colleges the same defects were discovered as in the high school: the inadequacy or non-existence of guidance and advisory systems, the maladjustment of students, the lack of flexibility due to an emphasis on credits instead of a program of continuous education. A large number of the students who had been tested in the last year of their high school course were again tested in the forty or more colleges in which

they were distributed at the end of their sophomore year. The purpose underlying the test was to discover what the bachelor's degree amounts to "in terms, first, of clear, available, important ideas, and, second, of ability to discriminate exactly among ideas and to use them accurately in thinking," objects which "are among the main reasons for acquiring an education and for the existence of educational institutions." An intelligence test was added in the sophomore examination. The tests included: (1) An intelligence test. (2) A "general culture" test, "ranging from very simple to very difficult questions, over the following fields: general science, 292 questions; foreign literature, 333; fine arts, 251; and general history and social studies, 346,"—a total of 1,222 questions. The questions were not based on the content of organized college courses but were intended to be "a fair measure of the permanent increment, the effective accumulations, attributable to a student's desire really to assimilate the ideas that constitute an academic education as contrasted with the urge merely to possess a degree as the result of having secured credits in a sufficient number of semester courses." (3) Two tests in subjects ordinarily included in formal college courses: English, 450 questions; mathematics, 210. Six of the forty colleges administered this test to students in all the four college classes.¹

It is unnecessary here to discuss the results of this test in detail. For purposes of the present discussion the significance of the test lies in its use as a method of assessing the growth of students. Comparable data which were made available in those colleges in which all the students took the test revealed the fact that in certain colleges there was no substantial difference between the four contemporary college classes in command of "enduring knowledge." Students can apparently obtain their degrees if they can accumulate a sufficient number of credits on the basis of examinations passed at the end of each semester course; the test revealed no evidence of a requirement that a student digest or retain knowledge which he could have acquired either in his courses or from other sources of education.

The same tests were given to these same students when they completed their senior year in 1932 and to high school seniors in 1933 and 1934. It was found that 85 per cent of the college students gained in varying amounts during their last two years in college. Details regarding these gains have not yet been published. But the most conspicuous fact revealed was variability, between institutions, among students taking the same courses in each institution, and in each institution among students taking different courses such as arts, science, business, education, and engineering. Further, there was a con-

¹ Carnegie Foundation for the Advancement of Teaching, *Twenty-fifth Annual Report*, pp. 51 ff.

siderable amount of overlapping among the high school seniors, college sophomores, and college seniors so that a percentage of college seniors were found to rank lower than a percentage of high school seniors. A system such as this, it was concluded, is defective in a large number of ways but primarily in attempting to deal with bodies of students as homogeneous when in fact they are highly heterogeneous, and, on the specious plea for equality, undertaking to provide for them varieties of courses and curricula which are assumed to be equivalent. As pointed out by Dr. Learned,

The net reaction of these penetrating solvents of old conditions has been to make a solemn emphasis on any particular external curriculum, as a means of educational salvation, seem futile and irrelevant. The important question in an individual's education has shifted entirely. Instead of asking solely, "Have you at some time learned this or that?" specified by the institution as of old, the question which it now seems most profitable to hold before each student is "What do you know?"—meaning by that "What knowledge, skills, and appreciations from the many sources available to you, have you, by virtue of your own evolving interests and abilities, gathered, organized, and matured within you to the point of effective mastery and use? You are an artist; your materials are all about you; what sort of educated man do you have in mind to seek to become?"¹

To put such an ideal before the student means not the prescription of a certain number of courses to be completed within a particular period of time; it involves on the part of those who are to guide and advise him a thorough knowledge of the goals attainable by him so that he may be placed in the way of his own self-education. But the progress of such self-education unless it is to be in a void must be subject to measurement, appraisal, and record, and for this purpose the instruments have now been developed. The existence of individual differences and of variability has been known for a long time, but little has been done to formulate types of educational program which would challenge the abilities of each to their fullest limits. What Dr. Learned calls the economics of self-education point to the same conclusion:

To invite or to compel every child to climb a uniform educational ladder to the bachelor's degree, even though doing so means moving the ladder downward while the student remains stationary, swells the expenditure in an inverted relation to the results, and achieves results that have little inner meaning when secured.

Rather let us start with a truthful and significant tested inventory, in terms of known standards, of what each child, when offered every reason-

¹ Carnegie Foundation for the Advancement of Teaching, *Twentieth Annual Report*, p. 60 (New York, 1935)

able aid and opportunity, progressively becomes through those powers of imagination and self-education that alone are potent. Such inventories, kept constantly open to correction, shortly project in the main both the quantitative and the qualitative aspects of the later educational problem for which the community should prepare.¹

The adoption of such measures for the distribution of education is justified further by social policy which demands not only a body of trained leaders but also a body of citizens appreciative of and able to understand "the true expert in ideas."

The testing program of the Pennsylvania Study has been criticized, as have other similar programs which have employed the new type, objective tests, on the ground that they test "mere knowledge" and fail to measure the power of organization and orderly expression, reasoning power, literary or aesthetic appreciation, or the other "higher processes," all of which, it is admitted, constitute the essential marks of a liberal education. The answer of those who have conducted the examinations is that "experience shows that the higher processes are directly conditioned by the comprehensiveness of important knowledge and by the clearness with which the ideas or the elements in such processes are held." Knowledge, in other words, is essential to thinking, despite the claims of those who minimize the value of knowledge on the professed ground that the purpose of education is to teach pupils "not what to think but how to think," or "to develop their power of reasoning."

Some were inclined to dismiss any test requiring a "yes" or "no" answer to disconnected statements, or even a choice among several opposites, as a test of "mere" knowledge without apparently realizing that knowledge is the necessary basis of all thought, and that if it be accurate and comprehensive there is at least strong reason to believe that the power to think must also be accurate and substantial in order to acquire and maintain it. The student who, from an adequate knowledge, can deal successfully with a hundred statements like the following: "The problem of capital versus labor had no place in the guild system of early modern times," must apparently be granted the power to think in that field from which the statements are drawn.²

The criticism has been met in another form by Dean Max McConn.

It comes to this: a high score in this kind of test does not infallibly demonstrate the attainment of what we call a liberal education; but a low score does infallibly demonstrate a lack of liberal education, because it re-

¹ *Ibid.*, p. 70.

² Carnegie Foundation for the Advancement of Teaching, *Twenty-fifth Annual Report*, p. 53.

veals the absence of the foundation upon which a liberal education must stand. Let me vary the figure. One may have a flourishing tree without fruit, but one cannot have fruit without a tree; knowledge—ample and accurate knowledge—is the tree on which the fruit we call culture must grow.¹

Of the many striking conclusions and implications of the Pennsylvania Study, not the least significant of which is the fundamental criticism of the discontinuity of the educational process resulting in American high schools and colleges from the artificial system of units and credits, the following are the most pertinent to the present discussion: (1) Learning students is a prerequisite to teaching them, or, to use a phrase which is now frequently used, education must be individualized. (2) Education must be regarded as a continuous process and must be based upon all kinds of information that can be cumulatively recorded about a student. (3) The abilities and capacities of students must be identified as early as possible and more adequate provision must be made for differentiated treatment of students in the assignment of programs of study. And (4) an integrated education demands early identification of needs, closer coöperation among teachers at different levels of education, and more coöordinated procedures, all of which can be made effective through the use of cumulative record cards which follow a student through his educational career.²

CUMULATIVE RECORDS AND EDUCATIONAL GUIDANCE

To conclude that the movement which has been described up to this point is merely a movement to substitute the new type or objective tests for the traditional form of examinations would be to place a wrong interpretation on it. Those who can speak with authority insist that the new type tests can measure only one but a very important aspect of the total process of education; they are equally insistent that other forms of examination—the oral

¹ Max McConn, "How Much Do College Students Learn?," *North American Review*, November, 1931, p. 451.

² On the Pennsylvania Study see Carnegie Foundation for the Advancement of Teaching, *Program for a Study of the Relation of Secondary and Higher Education in Pennsylvania, Progress Reports*, I-IV; and *Annual Reports*, *Twenty-fifth*, pp. 51 ff.; *Twenty-sixth*, pp. 31 ff.; *Twenty-seventh*, pp. 65 ff.; *Twenty-eighth*, pp. 39 ff.; and *Thirtieth*, pp. 47 ff. See also the following articles: W. S. Learned, "Credits' versus Education" in *Proceedings of the Associated Academic Principals of the State of New York*, 1933, and "Testing for Values in Education" in *Bulletin of the Association of American Colleges*, Vol. XX, No. 1, March, 1934; and Max McConn, "The Carnegie Foundation's Study of Secondary and Higher Education in Pennsylvania," in *Bulletin of the American Association of Collegiate Registrars*, Vol. 5, No. 2, and "How Much Do College Students Learn?," in *North American Review*, November, 1931, pp. 446 ff.

and the essay types—also have their place. What is claimed for the tests is that they offer a more accurate, more objective, more reliable and comparable measure of a student's command of the field of study than can be obtained by the traditional examination alone. Running through all the investigations which have been discussed is the implicit recognition that objective measures do not have the deterministic and predictive value that was at one time claimed by some of their too enthusiastic advocates. It is accordingly admitted that the use of one test alone or even of a number of tests at the same time have little greater value in contributing to the knowledge about an individual than the traditional examination. The development of and progress in the use of objective tests, if correctly interpreted, point to a new concept of education, that it is the responsibility of educators to discover the kind of education from which an individual may be expected to profit—to provide, in other words, the right education for the right pupil under the right teacher. This means not merely an appraisal of a pupil's intellectual status as measured by an examination, but the collection of as complete information as possible about the intellectual, personal, and social characteristics of an individual both in and out of school. If the central function of educational administration is the guidance of the individual to that kind of education which is most appropriate and adequate to his needs, interests, and abilities, such information must be both comprehensive and continuous. What is clear from the use of tests is that there is no single measure for predicting educational success.

CUMULATIVE RECORD CARDS

One of the most important outcomes of the testing movement has been the advocacy of the cumulative record card. Wherever efforts were inaugurated to discover information about the ability of pupils by means of tests given at a single time, as for example, in the last year of the high school course, it was found that the tests must be pushed back earlier until it is proposed to administer such tests throughout a pupil's educational career. As information derived from tests was accumulated, it was equally recognized that still more must be known about the health and physical condition, the social and economic background, the character and everything that goes to make up an individual's personality. Such information to be of value must be secured over a period of years and recorded on a blank which lends itself to easy interpretation.

The movement for the introduction and use of cumulative record cards was stimulated by a Conference on Vocational Guidance in Colleges called

in 1925 by the Division of Anthropology of the National Research Council, and was taken up by the American Council on Education. As a result of a conference convened by the Council in 1927 a Sub-Committee on Personal Record Cards was appointed; Secondary School and College Cumulative Record Forms, drafted for the Sub-Committee by Professor Ben D. Wood, were published a year later and have since served as a model for similar cards at the elementary school, high school, and college levels. The model record card was based on the following criteria:

1. The record form must show *trends of development* of abilities and interests.
2. It must be based on *accurate measures* and *concrete observations*.
3. The record must provide a means for recording measures and observations in comparable and meaningful terms, wherever such measures are available, but must at the same time provide for convenient recording and clear differentiation of whatever measures, subjective and non-comparable, may be available.
4. The data should appear in a form and order capable of showing their inter-relations, and thus presenting a coherent and integrated picture of the individual.
5. The record should be capable of quick reading; hence it should be in graphic form in so far as possible.
6. The record should be fairly complete for the large mass of "normal" children, requiring auxiliary cards only for extremely atypical subjects, mentally or physically.
7. The record should be reproducible, inexpensively, accurately and quickly, such as by photostating.
8. The record should be accompanied by a carefully written and amply illustrated manual of directions.
9. It should be administratively convenient, showing all available information on one continuous record form and permitting the collection of further data, by auxiliary cards and otherwise, for current use (in connection with the previous record) and for periodic sifting and entering on the permanent record.
10. Since all officers of the school that have to deal with students should have access to all the information that is available on each student anywhere in the school, it follows that the principal's record and the teacher's record should be duplicates so far as information of permanent significance is concerned.¹

The Secondary School Cumulative Record Form provides for the recording of information on the following items: Name, religion, sex, date of birth;

¹ From "Personnel Methods," *Educational Record*, Supplement, July, 1928, p. 17.

mental age, chronological age; intelligence quotients, school grade achieved; school grade attended; educational quotients; achievement test and school marks; height and weight; photograph; schools attended; record of attendances and absences; causes of absence, discipline; unusual accomplishments; mental, emotional, and physical experiences; extra-curricular experiences, athletic and non-athletic; clubs and offices; vocational experiences; educational plans; educational recommendations; vocational and professional preferences; interests reported; special defects; health; mental hygiene; social adjustments and home conditions; personality ratings; and measurements. Information on these items is recorded for each calendar year in such a way that the progress of the pupil can be traced easily across the record form.

EDUCATIONAL GUIDANCE

The value of such information when recorded in full detail and continuously over a period of years lies in the use to which it can be put for purposes of educational guidance. As contrasted with the snapshot or spasmodic information which may be derived from a single examination or test, the cumulative record form presents a moving-picture of the individual to be advised. The purpose of educational guidance is, at the risk of repetition, to discover the needs, interests, and abilities of a pupil or student and to provide for him that education from which he is most capable of profiting, an aim which is desirable as much from the point of view of the individual concerned as from that of the society in which he is a member. Educational guidance is based on the theory that it is harmful to force pupils to follow courses for which they do not have the necessary ability, a practice which may result in despair, a feeling of inferiority, habits of dependence, a smattering of knowledge, and superficial learning.

On the positive side educational guidance points to the need of greater flexibility and more differentiation of curricula and courses; it does not mean adaptation of education to the whims and caprices of the pupil, but an adaptation based on a knowledge, as accurate as present techniques may discover, of what a pupil can do.¹

The fundamental consideration which is involved here is whether a pupil shall be foredoomed to failure or whether, with such information as may be made available, he should be guided to make the most of his interests and

¹ The advantage of having available accurate, comprehensive, and continuous records in consultations with parents, who according to the headmaster of an English Public School are more difficult to handle than pupils or teachers, is not to be lightly set aside. Such records may be more successful than generalized statements in making parents see their offspring as others see them.

capacities.¹ There is still another consideration whether pupils shall be coerced to work under pressure of imminent examination or whether they shall be put in the way of becoming educated in a normal, unforced, and significant way by the provision of a curriculum appropriate to their needs and abilities. If the latter, then cumulative, comprehensive, and continuous information about each pupil as an individual is urgently to be desired, information which reveals his intellectual strength and weakness, his physical power or handicaps, his emotional qualities, and his personality as a whole. Educational guidance in accordance with this theory has now been made possible by the development of achievement and aptitude tests and the use of cumulative record cards. The latter have been compared by Dean Max McConn with the medical history and hospital chart, the former with blood counts and bacteriological tests. And just as no doctor will be satisfied with a diagnosis on the basis of a single test, so the educator, who is to guide and advise, must have available as comprehensive information as possible about a pupil.

The fear has been expressed that to adapt the curriculum to the revealed powers of the pupil would result in lowering of standards. On the contrary, it may be argued that the attempt to submit all pupils to the same curriculum has certainly had that effect in the United States.² To adapt the curriculum to the abilities and aptitudes of pupils will in fact make possible the attainment of higher standards in each field of work.³ What is sought in the United States is a way out of the impasse created by the constantly increasing enrollments in secondary schools and colleges. The American ideal of equality of opportunity, it has at last been discovered, cannot be realized in terms of mass education but rather by the individualization of education. Nor do marks obtained in courses provide an adequate basis for the differentiation of education which is implied, although such subjective judgments have their place. The situation is well summarized in a statement by Professor Edgar W. Knight of the University of North Carolina:

In time we may have to discontinue the present emphasis upon marks

¹ In contrast to this may be cited an example from an English secondary school in which the teacher informed a whole class at the beginning of the year that he had never known a pupil in that class to pass the external examination for which they were ostensibly to be prepared. Similarly in French secondary schools the complaint is frequently heard that there are too many *non-valeurs*. And yet such pupils, foredoomed to failure because of the rigidity of the curriculum, will in a few years be expected to assume the responsibilities of citizens alongside of those who will have "passed."

² See I. L. Kandel, *The Dilemma of Democracy* (Cambridge, Mass., 1933), and John L. Tildsley, *The Mounting Waste of the American Secondary School* (Cambridge, Mass., 1936).

³ Many years ago it was feared that the abolition of Latin and Greek as compulsory requirements at Edinburgh University would lead to a decline of classical learning. Actually it was found that the standards were raised because the classical course was taken by those who had a genuine interest and aptitude for it.

that are now given at the end of each course and that count for almost anything, and state requirements in different terms. As means of measuring the completion of requirements for the first two years [of college], the examinations must be intelligently designed. They may include any kind of tests, investigations, problems, or creative work by which the abilities and the performances of the students may be measured. Provision should also be made for a periodic, perhaps monthly, report by the instructor of each of his students, a statement setting forth his observations of the student's habits, health, character, personality, and any traits or qualities, which in the opinion of the instructor may influence the effectiveness of the student's work. By these means a student's past achievements and present interests can be measured more nearly and accurately and his future needs can be more nearly determined than at present.¹

It is interesting to note that the International Commission of the New Education Fellowship considers the cumulative record card as "the way out" of the examination tangle.² Among the few examples of schools or school systems which have adopted such cards is the Education Committee of Kent, England, which has introduced the Pupil's Record Card in its elementary schools. Among the uses for which it has been designed the Record Card is (1) to serve as an indication of the level of work to be expected from a pupil; (2) as a first step in diagnosis; (3) for the information of new teachers in cases of change of class or school; (4) for changes in class and school organization. This analysis concludes with the statement that

The utility of the Record Card will be greatly extended when it is possible to make more use of it in deciding the course of post-primary education most suited to a child's needs. It seems reasonable to suppose that, ultimately, a scientific, objective, and normative record of a child's school career extending over many years will come to be recognized as a safer guide in this than a one day examination, however skilfully conducted. The data on the Record Card (particularly the I.Q. and E.R.) should be quoted on behalf of pupils who are candidates for Special Places in Secondary Schools or when the transfer of a child from a Central to a Secondary, Technical or Art School is being recommended. Finally, the Record Card does not cease to be of use at the end of a child's school career or even after he has left school. It gives information valuable to the person who seeks to place the child in the employment most suited to his capacities, whether this person be the Head Teacher or the Vocational Guidance Officer.³

¹ E. W. Knight, *High School Journal*, November, 1934, pp. 237 f.

² *The Examination Tangle and the Way Out*, Part III (London, 1935).

³ Kent Education Committee, *Notes on the Record Card*, p. 3 (Maidstone, 1934). See also *Education in Kent, 1928-1933*, pp. 49 ff. (Maidstone, 1934).

It is impossible to say how widespread is the use of the cumulative record card, whether in the forms issued by the American Council of Education or in other forms.¹ The use of such cards has been strongly urged on and widely adopted by the independent schools associated with the Educational Records Bureau. They were employed effectively in the Pennsylvania Study. Two examples may be cited to indicate both the type of information gathered and the use to which it may be put.

*The Case of Margaret Elinor Crawford.*² The record covers the school life of this child from the age of eleven to seventeen (line 4)* beginning in the second half of the sixth grade (line 3) which she entered in February (line 5), 1925 (line 2), at the Howard Elementary School (line 26).

The grades given by Margaret's teachers in earlier years are summarized on the left of the score frame. Her spelling is rated A, music and drawing, C, and all other subjects B (cols. 1-2). A glance at her record as a whole shows that her teachers consider her a remarkably able pupil. Except for a C in writing at the end of the second semester, 1925 (col. 4), there are no grades below B during the entire six years and A ratings are frequent. Her standard test scores are heavily concentrated in the highest part of the card and in general run higher than the teacher's grades. There is none below the 83rd percentile; eight only are below the 93rd percentile, and thirty-seven are above the 98th percentile. This means that in the latter subjects, more than 98% of the children of like age taking these tests scored below Margaret.

Details of the standard test scores are to be found on the reverse side of the card. For the sake of clarity these are much abbreviated on the chart. In May, 1925, 1926, and 1928, and in March of 1927, Margaret tested consistently in the 99th percentile in intelligence. The other tests of these years consisted of the Stanford Achievement tests—a group of ten tests in various subjects each having a separate score in addition to the score for the group. The latter is recorded as the "Stanford Achievement" score ("St. Ach."). In Margaret's case the sections of the test spread slightly downward after the first year due to the increasing selection of the group on which the test is standardized and possibly also to the fact that the pupil is not doing her best because of constant association with inferior pupils. Since 1925 she has been ready for 10th grade work although compelled to sit through the 7th, 8th, and 9th grades and to work at their

*NOTE: The lines of the Record are numbered on the extreme left and the columns in line 5

¹ The American Council's form has been adapted for use in elementary schools. See Margaret W. Moore, *Cumulative Educational Record for Elementary Schools*, prepared for the Cooperative Test Service (Washington, D. C., n.d.).

² The Carnegie Foundation for the Advancement of Teaching, *Study of the Relations of Secondary and Higher Education in Pennsylvania*, Progress Report IV, pp. 3 ff.

1		Crawford		Margaret Elmer		4-21-14 Penna		M O B O		SEX		COLOR					
2		YEAR		1925		1926		1927		1928		1929		1930		1931	
3		GRADE		6		7		8		9		10		11		12	
4		AGE		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
5		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
6		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
7		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
8		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
9		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
10		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
11		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
12		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
13		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
14		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
15		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
16		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
17		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
18		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
19		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
20		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
21		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
22		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
23		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
24		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
25		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
26		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
27		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
28		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
29		MIND		10-13-27		10-12-6		10-13-5		10-13-2		10-13-2		10-13-2		10-13-2	
30		MIND		10-13-27		10-12-6		10-13-5		10-13-2							

STUDENTS ADDRESSES		H = HOME S = SCHOOL T = TELEPHONE	
54			
55	2		Initiative
56	1		Initiative
57	0		Responsibility
58	-1		Charm
59	-2		Personal Appearance
PERSONALITY MEASUREMENTS			
60 NAMES OF TESTS USED			
NAME		HEALTH	RELIGION
61 FATHER		Good	
62 MOTHER		Good	
63 STEP PARENT			
64 (GUARDIAN)			
65 A		LIVING AT HOME	SCHOOLING
B		Occupation	AGE
C		EDUCATION DEGREE AND KIND	OCCUPATIONS
D		ADDRESSES	TCL
E		LANGUAGE SPOKEN IN THE HOME	BEFORE 10 AFTER 10
F		TYPE OF HOME COMMUNITY	BEFORE 10 AFTER 10
G		IF PARENTS ARE SEPARATED GIVE DATE	

Notes

66 YEAR	1925	1926	1927	1928	1929	1930	1931
67 AGE	11	12	13	14	15	16	17
68							
69		Good	Good	Good	Good	Good	Good
70		Good	Good	Good	Good	Good	Good
71		Good	Good	Good	Good	Good	Good
72		Good	Good	Good	Good	Good	Good
73		Good	Good	Good	Good	Good	Good
74		Good	Good	Good	Good	Good	Good
75		Good	Good	Good	Good	Good	Good
76		Good	Good	Good	Good	Good	Good
77		Good	Good	Good	Good	Good	Good
78		Good	Good	Good	Good	Good	Good
79		Good	Good	Good	Good	Good	Good
80		Good	Good	Good	Good	Good	Good
81		Good	Good	Good	Good	Good	Good
82		Good	Good	Good	Good	Good	Good
83		Good	Good	Good	Good	Good	Good
84		Good	Good	Good	Good	Good	Good
85		Good	Good	Good	Good	Good	Good
86		Good	Good	Good	Good	Good	Good
87		Good	Good	Good	Good	Good	Good
88		Good	Good	Good	Good	Good	Good
89		Good	Good	Good	Good	Good	Good
90		Good	Good	Good	Good	Good	Good
91		Good	Good	Good	Good	Good	Good
92		Good	Good	Good	Good	Good	Good
93		Good	Good	Good	Good	Good	Good
94		Good	Good	Good	Good	Good	Good
95		Good	Good	Good	Good	Good	Good
96		Good	Good	Good	Good	Good	Good
97		Good	Good	Good	Good	Good	Good
98		Good	Good	Good	Good	Good	Good
99		Good	Good	Good	Good	Good	Good

pace. Tests in new material in May, 1928, (col. 16)—Cross English test, Brown-Woody Civics, Junior Latin and C.R.B. Algebra—place her in the 98th and 99th percentiles. These percentiles are based only on the local groups and are therefore indicated by crossed open circles $\oplus$.

In the senior high school Margaret tests about equally well in all types of work. She took general science throughout the junior and senior years, and in May, 1929, took a special test prepared for college freshmen and scored in the 91st percentile. Her interest in science led her to do a large amount of supplementary reading that summer, and when she returned in October she repeated the test and raised her score to the 98th percentile (col. 22). This score she raised still higher in May, 1930 (col. 24). Thereafter she dropped science, but a year later, in 1931, she could still test to the 94th percentile in that subject (col. 28).

Latin was begun in the fall of 1927 and the Junior Latin test the following May showed Margaret at the head of the local group (col. 16), although the teacher gave her an A—. That this local score was a real achievement is shown by the standardized tests of 1929, 1930, and 1931—in all of which she gained the 99th percentile (cols. 19, 20, 24, and 28). In her junior year her Latin teacher suggested that she begin French by herself and gave her some preliminary lessons on the pronunciation. Margaret worked on grammar and easy reading during that summer and in the fall took the Columbia Research Bureau French test, scoring in the 83rd percentile. On the strength of this performance she entered second-year French and repeated the test twice during the school year, each time with noticeable improvement (cols. 26, 27, and 28).

It should be noted that in the latter part of the senior year Margaret had the opportunity in every field of her secondary study to demonstrate her attainment by means of comprehensive objective examinations. Through constant use as a part of her education these tests had become a matter of routine to her and had given her the attitude that knowledge once acquired was to be retained and deliberately improved for permanent use—not crammed for recitation and a term credit and then forgotten. It is apparent, too, that the school had given her encouragement in independent work by testing it and recording her achievement. Her special work in general science and in French have already been described. At the beginning of her junior year she applied for a test in ancient history in which she had long been interested, though she had taken little of it in school; the result in this test showed the same high quality as her other work (col. 18).

The personal history that parallels Margaret's intellectual development furnishes important items. Under notable accomplishments (line 31) are a prize essay in 1925, the reading of *Barrie* complete when twelve years of

age, and, when thirteen, the reading and reviewing of all Shakespeare's historical plays during a summer vacation. She was an active club girl, occasionally an executive (line 35). In her early vocational choices (line 47) teaching, medicine and law rotate. She finally decides on teaching, then on college teaching, then on teaching of Latin and French.

Data on the back of the card indicate that Margaret comes from a normal American family in which the parents are without higher education (lines 61, 62). She has four brothers, three of whom have been at college (line 65). Her personality ratings can only be regarded as vaguely suggestive. Judgments on this point are equivocal since they lack definition of the behavior on which they are based.

Altogether Margaret's picture is that of a remarkably capable, steady, and independent mind developing admirably with an excellent background. She is undoubtedly a college risk of the first class.

*The Case of John Morton Smith, Jr.*¹ The facts concerning this boy were accidentally discovered in the records of a large city school which took exhaustive measures of its pupils but made no use of them because of the lack of an effective way to assemble and present them.

In 1927 John took college entrance examinations in five subjects with results indicated by the open circle in that column. He was refused admission because in English he ranked among the lowest 16 per cent, because the principal, supervisor, and mathematics teacher (all strangers to him) estimated his intelligence as below average, and because in a three-minute interview the college admissions officer could make nothing out of an excessively shy, self-conscious, and excitable lad.

The unused data in possession of the school when brought together in cumulative form tell a very different story; in fact they seem to indicate that the college officer who rejected this boy was actually dealing with a mind that should have been classed among the best 5 per cent of college risks. There are ten measures of John's intelligence as shown in the I.Q. graph beginning in 1922, none of them falling below the ninety-fourth percentile. There are four standard measurements in geography, three in American history, and one in economics—all above the ninety-seventh percentile. English fluctuates more widely but is nowhere below the eighty-fifth percentile, while of seven measures in French all but one are in the highest 2 per cent of modern language ability for the respective ages. The science line—general science, biology, physics, and chemistry—is lower but still well above the average. The four arithmetic measures are good, but the symbolic thinking required in algebra is indifferently done, and geometry is far down. Fearful of failure in the college examination in

¹ Carnegie Foundation for the Advancement of Teaching, *Program for a Study of the Relations of Secondary and Higher Education*, pp. 2 ff. (New York, 1928).

Smith

John

Morton, Jr

LAST NAME

FIRST NAME

MIDDLE NAME

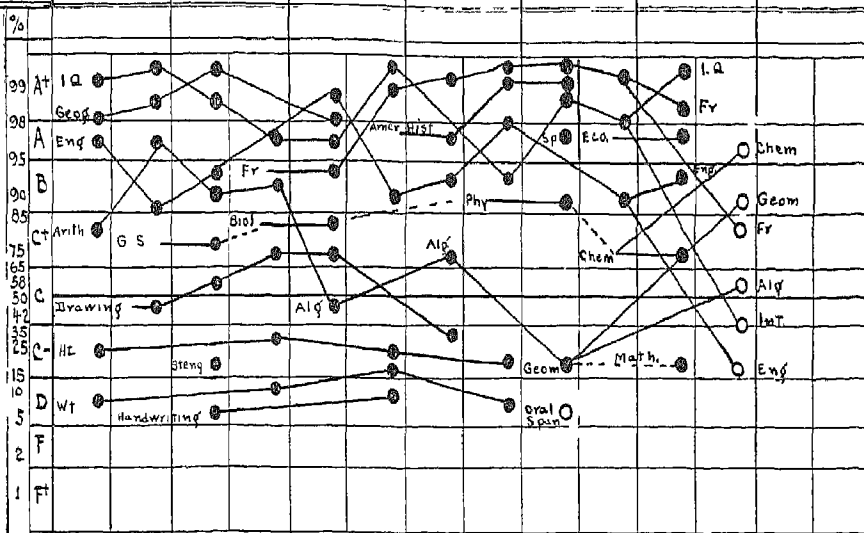
RELIGION Prot

(BOY) GIRL

DATE OF BIRTH March 15 1912

YEAR	1922	1923	1924	1925	1926	1927	1928
MENTAL AGE	13						
CHRON AGE	10	11	12	13	14	15	16
INTELLIGENCE QUOTIENTS	130	132	129 123	123 133	120	130 128	132
SCHOOL GRADE ACHIEVED	7	8	9	10+			
SCHOOL GRADE ATTENDED	6	7	8	9	10	11	12
EDUCATIONAL QUOTIENTS							

ACHIEVEMENT TEST AND SCHOOL MARKS
HEIGHT AND WEIGHT I Q'S
STANDARDIZED TEST
SUBJECTIVE MARKS



PHOTO

TESTS AND SCORES

SCHOOLS (PRIVATE OR PUB)	Payk J S H S
CITY	P
NO DAYS ATTENDED	
ABSENT	
CAUSE OF ABSENCE	
DISCIPLINE	
UNUSUAL ACCOMPLISHMENTS	Read Shakespeare complete - Essay on Shakespeare in politics
EXPERIENCES	Read complete Hugo, Balzac, Moliere, Racine in French in summer
MENTAL	Translated 3 French comedies into English
EMOTIONAL	Learned Spanish
PHYSICAL	Wrote book 200 pages on Geography in French literature

YEAR		1922	1923	1924	1925	1926	1927	1928
CHRON A		10	11	12	13	14	15	16
EXTRA CURRICULAR EXPERIENCES	ATHLETIC	H	B, H F, H	B, H Ten	B, H H, H	H, H H, H	H, H	
	HRS PER WEEK							
	NON-ATHLETIC	Dr	Dr Dr, Deb	Dr, Deb, J Dr, J	JJ	JJJ	JJJ	JJJ
	HRS PER WEEK							
CLUBS OFFICES	FRATERNITIES							
VOCATIONAL EXPERIENCES	TYPE	Bookstore messenger	Bookstore clerk	Law office boy	Newspaper office boy	Newspaper reporter		
	DURATION	3 mos. Summer	3 mos. Summer	3 mos. Summer	3 mos. Summer	3 mos. Summer		
	PAY PER WEEK	\$ 6 00	\$ 10 00 - 15 00	\$ 20 00	\$ 20 00	\$ 25 00		
	SUCCESS	++	+++	+++				
ED PLANS		L A	L A	L A	L A	L A		
ED RECOMMENDATIONS		L A	L A	L A	L A	L A		
VOC & PROF. PREFERENCES		Writer	Writer	Law	Journalism	Writer		
INTERESTS REPORTED		Eng Lit	Eng Lit	Fr Fr Lit	Hist	Sp Lit		
MEASURED								
SPECIAL DEFECTS		Astigm C						
HEALTH		-	-	+	+	+	+	
MENTAL HYGIENE								
SOCIAL ADJUSTMENTS AND HOME CONDITIONS		Very shy		Very shy	Avoids company	Avoids company		
PERSONALITY RATINGS	Pers	-2	Pers -1	Pers -2	Pers -2	Pers -2		
	Leadership	0	Leadership 1	Leadership -1	Leadership -1	Leadership +2		
	Initiative	-2	Industry +2		Initiative +2	Initiative +2		
			Persistence +2					
	Reliability	+2	Reliability 0	Reliability +2	Reliability +2	Sense of honor +2		
MEASUREMENTS			Introvert					

geometry, John's parents hired a tutor who had him commit to memory as many theorems as possible with the purely chance result there indicated. Drawing hovers about the average. A Stenquist test of mechanical ability shows poor success. Handwriting is very poor, thus doubtless explaining the low college entrance mark in English -the paper was illegible. Height and weight lines are much under average. Desiring to enter college a year earlier, the boy learned Spanish by himself when fourteen. The school refused to admit him to the college examination, but his teacher gave him a standard American Council Spanish test in which he scored in the ninety-seventh percentile for third-year students. The rating in oral Spanish at the same time was, of course, low.

In extra-curricular activities the scattered obtainable facts about this boy make an impressive showing when they are pieced together. When twelve years old he brought to his English teacher an essay on "Shakespeare in Politics" highly documented from nearly every one of Shakespeare's plays. The following summer, having studied French for two years, he read by himself certain editions of four French authors and took the French teacher's examination thereon. Later he translated three short French comedies into English, and in his last year presented a lengthy study of geography that he had made from his French readings. Such work as this apparently earned him the maximum rating of "initiative" at the bottom of the card, although his first rating in that trait is the minimum because it was given by those in charge of group activities which he disliked. His "personality" rating is uniformly low. A psychiatrist rated him in 1923 as markedly introverted. His athletic activities include a little baseball (B), football (F), and tennis (Ten) in early years, but after the age of thirteen these group sports give place to solitary hiking (H) with a book. Similarly, dramatics (Dr.) and debating (Deb.) give way to what is called "journalism" (J), which, however, was discovered to consist entirely of reviewing learned books for the local paper. His success at his summer jobs is good, his ambitions and interests are steady and consistent.

Studied as a whole, this record gives indubitable indications of a mentality that may be close to genius. The separate items taken alone count for little but the sweep of evidence across even six years of this boy's life is unmistakable and should place the subject among those whom an institution handles with the utmost care. It is doubtless an extreme instance, but, if a case so obvious as this can be so badly bungled in our administrative procedure, it seems likely that the average child must be a frequent sufferer.

After reviewing the different methods which have been tried to secure some reliable and accurate measures for predicting the probable success of

students, Professor Strang of Columbia University concludes with a statement which constitutes an appropriate summary of this chapter:

Predictions based on the data from cumulative records, involving as they do the many-sided aspects of the individual, seem to be the most satisfactory basis on which to advise boys and girls concerning their future educational plans. The cumulative record prevents a single non-typical performance from being given undue weight.¹

¹ Ruth Strang, *Personal Development and Guidance in College and Secondary School*, p. 133 (New York, 1931).

CHAPTER IV

EDUCATION, EXAMINATIONS, AND THE INDIVIDUAL

THE discussion in the preceding chapters began with the criticisms of the traditional system of examinations in the United States. After a detailed consideration of the substitutes for traditional examinations it closed with the conclusion that the problem of examinations is not primarily one of discovering more accurate scientific and technical methods of constructing and scoring examinations. The problem of examinations strikes at the very roots of the whole meaning and significance of education for society. It raises the question whether society can proceed, as it has in the main done in the past, on the assumption that there is only one pattern of culture or liberal education to which all must be moulded, as they advance from elementary to higher education. It involves a consideration of the meaning and effects of failure in education, and posits the question of the profit derived by a pupil exposed to standards of education, however sound and laudable they may be in theory, if he fails to come within measurable distance of attaining these standards.¹ It inevitably brings up the problem of the social and economic distribution of individuals in society and the danger to social stability which may result from educational and vocational maladjustment. In the present crisis it raises the issue whether by an appropriate organization of education the emergence of educated proletariates with the tragedies that have been witnessed in Europe as the direct results of overcrowding in certain professions can be avoided. This issue in turn leads to the question whether it is not the function of an educational system to promote the best happiness of the individual by putting him in the way of the highest development of which he is capable and which will contribute to the best progress of society itself.

To approach the problems of education in this way may be a shock to those who fear that further differentiation in education and adaptation to the interests and aptitudes of individual pupils and students may mean a departure from the Great Tradition, meaning by that the classical tradition, a knowledge of the best that has been said and thought in the past. While no one

¹ When the doctrine of formal discipline dominated education, it was claimed that even though these standards were not attained, the "mind" was trained on the way. With the reservations and limitations that have been placed upon this doctrine, the problem in education has become one of discovering such content as will provide intellectual training and at the same time have what Dr. Abraham Flexner described as "affirmative values" for the learner.

would be disposed to doubt the sincerity of those who have a continued and abiding faith in the Great Tradition, it may still be maintained that they ignore the pragmatic demands of the situation which is developing everywhere. Those who profited from the Great Tradition were in the past a selected minority; to-day the task is twofold: first, to re-define the concept of liberal education in terms more appropriate for the world in which we live, and, second, to provide appropriate types of education for the increasing numbers who are entering on some form of post-primary education and proceeding even beyond that stage. As these numbers increase, the range of individual differences of aptitude increases. It would be unfortunate for the world if the ideals of the Great Tradition and with it the ideals and standards of scholarship were allowed to disappear, but all the evidence that can be gathered from the study of examinations points to the fact that success in this type of education demands special aptitude. It may even be asserted that the selection of those who have the ability to profit from it would help to perpetuate and strengthen it.¹ Hence a demand for greater differentiation in programs of education, considered from one point of view, does not mean that the Great Tradition must be discarded but that it should be retained as one type in the program of differentiation.

Society is to-day confronted with a different task in the organization of education from that which it was called upon to meet when one type of education was provided for the masses and another for the small minority. Whatever the political organization may be, the importance for society of a body of well-trained individuals is everywhere being recognized. This recognition does not, however, mean the same education for all. It means that measures must be devised whereby each individual may receive that type of education from which he can best profit. The traditional system of selection by examinations is beginning to be found inadequate for the purpose, whether from the point of view of individual pupils or of society itself. The essence of the problem is not the reliability of examinations and their scoring, but the validity of education. Merely to discover the interests and abilities of individuals is not enough; there must also be another criterion and that can be found mainly in the social needs and requirements. Merely to follow individual interests and abilities in devising educational programs has too frequently been to follow the line of least resistance, to tread the primrose path. In education as in life freedom always demands as its correlative a sense of responsibility. And this responsibility is to be found in the common duties which all members of a society must fulfill. In the words of Vittorino da Feltre,

¹ See note on p. 144.

Not everyone is called to be a lawyer, a physician, a philosopher, to live in the public eye, nor has everyone outstanding gifts of natural capacity, but all of us are created for the life of social duty, all are responsible for the personal influence which goes forth from us.¹

The problems and questions which have been raised are not confined to the United States; they are to-day world-wide. The United States is, however, in a better position to make a contribution to their solution because of conditions which do not as yet prevail elsewhere. There is, first, the widespread acceptance of the ideal of equality of opportunity, which after long years is beginning to be rescued from the danger of confusion with identity of opportunity. Secondly, the adaptation of education to changing social needs is not subject to the coercive restraints of a traditional concept of culture, the possession of which is the key to certain social and economic privileges.² Thirdly, there exists in the United States a venturesome spirit of experimentation, which, though at times and in some hands subject to the crudities of the half-educated man, does over a period of years yield salutary results.³ Fourthly, the consideration of educational problems and action based upon it are not vested in the hands of official administrative authorities.

While all these factors may from one point of view be regarded as detrimental to anything like stability in educational ideals and practices, a study of the educational situations in other countries only reveals the existence of the same problems everywhere else and the urgent need of meeting them soon, if the unrest of the rising generation is to be allayed.⁴ The problems here discussed are as much social and economic as they are educational. Everywhere the need of providing more education at the secondary level is beginning to be recognized;⁵ in most countries overcrowding in the univer-

¹ W. H. Woodward, *Studies in Education during the Age of the Renaissance, 1400-1600*, pp. 12 f. (Cambridge University Press, 1906).

² It is not always clear in those countries where the tradition of culture has been most ardently preserved whether that has been possible because of a desire for education for its own sake or for the rewards that accompany its attainment. The traditional program has at the same time served as a method of selection. Thus the French educational authorities in order to curb overcrowding in the medical profession introduced the classical studies as a compulsory requirement for students of medicine.

³ It happens too frequently that though a press interested in exploiting the sensational the foreigner hears more of these crudities than of the solid accomplishments. But even in the case of an authoritative work the foreigner who derives his picture of higher education in the United States from Abraham Flexner's *Universities, American, English, German* (New York, 1930) has failed to read Dr. Flexner's prefatory statement (p. viii) that he has given more attention to a consideration of where universities now satisfy and where they fall short of his own "idea of a modern university," than to their merits.

⁴ See W. Kotschnig, *Education Dirigée, Etude Internationale sur le Surpeuplement des Universités et le Chômage des Etudiants Diplômés* (L'Entr'Aide Universitaire Internationale, Geneva, 1935).

In the United States, as already indicated in the preceding chapters, the problems reach into the college level.

sities is looked upon as a social menace; in many countries there is a search for methods and techniques by which the right education may be provided for the right pupil.

A study of examinations, unless it is approached in the light of the issues here raised, must inevitably concern itself merely with the machinery of marking in order to adhere to some arbitrary standard.¹ But a candidate or the teacher of a candidate who has passed or failed knows little more about what the examination has revealed concerning his ability. Nor is it very helpful to the teacher to learn from the blanket reports on examinations that

In the set books of the English paper the more mechanical part of the work (i.e. contexts, reproductions, etc.) was usually adequate, but questions which called for any originality of treatment too often revealed a lack of resourcefulness. . . . In history the general standard reached seemed to be somewhat lower than usual, though in some schools the answers to questions on social and economic history had definitely improved. Insufficient attention was paid to geography in most schools. . . . In the chemistry section of the physics and chemistry paper, the general standard was definitely weaker than in recent years, many candidates being largely ignorant of the subject. Answers failed, however, not only through ignorance of facts, but because of an outlook on the subject wholly devoid of common sense.²

Such a report no doubt has its value but it tends to perpetuate a game between teachers and pupils, on the one hand, and examiners on the other; neither the general statements nor the individual results give the teachers much guidance on how to deal effectively with each pupil. The most frequent objection which is brought against the new type, objective test is that its measures "mere knowledge" only and not the higher processes, which it is claimed the traditional essay examination can test adequately. The following statement by the Panel of Investigators appointed to inquire into secondary school examinations in England is pertinent on this point:

Admittedly Examiners are oppressed by the fear that candidates may obtain pass or credit by 'cramming,' a process which is not always distinguishable from careful teaching and intensive work. To prevent cram-

¹ "For each paper the crucial question to be answered was 'What in terms of performance, does a pass or credit in this subject mean?' or, more particularly, 'Did the scripts of the candidates who just passed reveal work which was tolerable or was it in fact of little or no worth?' 'Could the work of the candidates who just got credit be regarded as reasonably good?' " From *The School Certificate Examination being the Report of the Panel of Investigators appointed by the Secondary Schools Examinations Council to Enquire into the Eight Approved School Certificate Examinations*, p. 55 (London, 1932).

² From the report of the Oxford and Cambridge Schools Examination Board, presented to Convocation, quoted in *The Manchester Guardian*, February 11, 1936.

ming, questions are set inviting candidates to contrast, compare, or criticize groups of facts or individuals. But the experienced teacher is equal to this device, and the pupil with a good memory absorbs the subheadings of a comparison or criticism from dictated notes as easily as he crams a series of facts from a textbook. The pupil with a weaker memory writes down such facts as seem more or less germane to the question. In the latter case success or failure depends on the Examiner's attitude; some consider an accurate knowledge of such facts worthy of a credit; others hold that this knowledge, unless logically applied, scarcely deserves a pass. Examiners report regretfully that such questions either produce irrelevant answers or attract very few candidates. In the judgment of the Investigators the sustained attempt to make History papers test the logical and critical powers of the sixteen year old candidate has proved a failure.

Perhaps even less satisfactory results have ensued from the use of 'academic' or abstract language. Most pupils who take the examination can understand concrete facts and attach the right meaning to straightforward questions. Few have the knack of absorbing generalities or discussing them with plausibility. The results of training an immature mind to deal in ideas and idioms beyond its grasp are short-lived, so far as they concern historical knowledge. But there may be permanent effects on a candidate's intellectual sincerity.¹

This may be supplemented with a further statement from the same report:

The mass of scripts which have been placed before the Investigators affords convincing evidence that by encouraging the reproduction of lifeless textbook formulae the existing type of examination deadens the pupil's interest in what should be one of the most stimulating subjects [history] in the school curriculum.²

The reports of the examiners of the University of Paris on the *baccalauréat* examination of 1935 are of a similar character.³ In the natural sciences the examiners found that

The defects most frequently noted by the examiners, apart from lack of information shown by some of the lazy or unprepared candidates, was the deficiency in method of exposition, absence of plan, negligibility of drawing, and carelessness in grammar and spelling. In the oral examination, while there are candidates who give proof of intelligence and thinking, there are still too many who yield to the tendency of trying to present

¹ *The School Certificate Examination*, p. 81.

² *Ibid.*, p. 84.

³ The quotations which follow are cited from a mimeographed circular, *Rapport sur les Examens du Baccalauréat de l'Enseignement Secondaire en 1935*, issued by the *Office du Baccalauréat* of the University of Paris.

an entirely formal exposition without order or logic. It is difficult to make such candidates give precise facts or even to get them to consider the observations and experiments which serve as the basis of their conclusions or theories.

The examiners of both the written and oral examinations in French are unanimous in directing the attention of candidates to the importance of writing correctly and with due regard to spelling. 'Of 134 papers,' writes one examiner, 'I hardly found 20 which were free from grammatical mistakes and errors in spelling.' Further, too many candidates showed themselves incapable of drawing up a plan. 'One has the impression,' according to one report, 'that the pupils start right off without thinking and without a plan.' Frequently some candidates 'instead of dealing with the topics set develop topics already prepared.'

In the modern foreign language papers

a consideration of the marks obtained in the written and oral examinations, in lieu of detailed reports by the examiners, shows that while a sufficiently large group of candidates have given evidence of adequately sound mastery, the rest, a more numerous group, do not appear to have derived much advantage from the instruction which they received in the six years of the course.

According to the report of one examiner in history and geography the candidates in general seem to have a sufficient knowledge of the course. As usual too many have obviously used *résumés* and it is useless to expect them to reflect on the facts and to indicate the relations which exist between them.

As in the case of the examinations conducted by the New York State Board of Regents and the College Entrance Examination Board the retention of the external examinations in England is justified on the ground that there is almost perfect agreement between the results and the teachers' expectations.¹ Thus the Secretary of Local Examinations, Oxford University, at the close of a critique of *An Examination of Examinations* conducted by Sir Philip Hartog and Dr. E. C. Rhodes, finally appeals to the teachers

Perhaps I may be allowed to send a word of comfort and advice to young teachers who have not yet joined the conspiracy of silence and whose consciences are still tender. They must be feeling that they will never again dare to mark their pupils' work or say that an answer to a question is

¹ One examiner in the report on the *baccalauréat* mentioned above states that the school records (*livrets scolaires*) presented by candidates from *lycées* and *collèges* in general reveal them standing accurately; according to two examiners the same is not true of records from private schools.

good, bad, or indifferent. They may take heart, for they will find that, somehow or other, the verdicts of examiners coincide with their own in a very large number of cases and that this number will increase as they gain experience. In such cases they can use the examination results to vindicate their own forecasts.¹

Such a justification fails to meet the educational implications involved. For if teachers can forecast the results of examinations so successfully, it would add to their professional status and dignity to be allowed to certificate students themselves, since the margin of error is so small. And, secondly, if pupils can succeed only to the degree indicated in the reports of the Oxford and Cambridge Schools Examination Board and the Panel of Investigators, then the efforts to secure reliability of marking are not justified by the results, and the teachers should be placed in the position of making constant appraisals which they can interpret adequately and whose interpretations will enable them to deal with each pupil appropriately. The justification of an educational system does not rest upon ability to anticipate success or failure and to prove these anticipations by examinations, but rather in using such information to adapt the work to the known abilities of the pupils.

This, it is submitted, is what is claimed for the new type tests, not that they can be more accurately and more reliably scored but that they furnish a more adequate basis for guiding, helping, and advising each pupil. It is admitted by those who can speak with authority that the tests themselves are not yet perfect. The same admission has been made both by the Secretary of the College Entrance Examination Board in the United States (see p 48) and by the Panel of Investigators in England (see pp. 154 f.) But another claim is also made as the result of experimentation over nearly two decades, that no test, however perfect, can tell everything that needs to be known about a pupil.

If it is true, and there is now adequate reason to believe that it is true, that so far as the machinery of scoring is concerned the results obtained from the new type test are more reliable than those secured by the traditional examination, and the former have the further advantage of providing diagnostic data, the situation becomes more aggravated when the traditional examination—the spasmodic, snapshot picture of a candidate—is used not only to

¹ *The Times Educational Supplement*, December 21, 1935, p. 116. In a later article, however, in the *Journal of Education* ("Examinations"), May 1, 1936, pp. 280 ff., Mr. Burnet does not seem to be inclined to show the same unlimited confidence in the school records which are sometimes submitted. The Northern Universities Joint Matriculation Board, to meet the criticisms of examinations implied in the Hartog report, has also published a report describing the methods of marking papers, which prove that teachers' expectations are fulfilled in 92.44 per cent of the cases.

measure what he has achieved but what his promise for further progress may be, that is to say, for purposes of selection for a particular type of education. In England Professor C. W. Valentine of the University of Birmingham has shown the unreliability of the first selective examination by which boys and girls at about the age of eleven are sorted out for secondary, central, and "senior" schools.¹ In the first secondary schools examination, which is both a terminal and a qualifying examination, the average percentage of passes over a period of years has been 65 per cent with the probability that the examination is not taken by all the pupils who have spent four years in a secondary school.

In France a generous system of scholarships for post-primary education is administered on the basis of a traditional examination; fees have been abolished in the secondary schools but the authorities are still struggling with the problem of devising a reliable method for selecting those who are most likely to succeed in the type of education there offered. In the *baccalauréat* examinations at the close of the secondary school career only 45 per cent of the candidates pass the first part, and of these about 55 per cent pass the second part, while the distribution of passes in the various academies, administrative areas for education, ranges from 20 per cent to 80 per cent.

In Germany an investigation sponsored by the former Minister of Education in Saxony, Dr. W. Hartnacke,² revealed the diversity of standards in the secondary schools of the state, which in the past have always been regarded as homogeneous in standards, and the diversity of abilities among the students who have nevertheless crowded into the universities as a supposedly homogeneous group. Admissions to the German universities have been reduced to about one-third of the number entitled to proceed from the secondary schools on the basis of the *Reifeprüfung*,³ but no provision has yet been made in the form of differentiated types of education for those adolescents who in future will hesitate to undertake the hazard of a nine-year course, only to be turned aside from the universities in the end.

Examples of the same kind could be cited from other countries.⁴ The con-

¹ C. W. Valentine, *The Reliability of Examinations* (London, 1933) The "senior" school is the recently developed upper section of the elementary school for children from eleven up to fourteen or fifteen.

² W. Hartnacke and E. Wohlfahrt, *Geist und Torheit auf Primanerbanken* (Dresden, 1934).

³ For the Law against Overcrowding in the Universities (April 25, 1933) and the subsequent decree of December 28, 1933, see I. L. Kandel, *The Making of Nazis*, pp. 102 ff. (New York, 1935). Since this reduction has been in effect a smaller number than those actually entitled to do so under the regulations have actually proceeded to the universities.

⁴ The three Scandinavian countries, Norway, Sweden, and Denmark, are confronted with the same problem: overcrowding in the secondary schools, leading to overcrowding in the universities, leading in turn to overcrowding in the liberal professions, and accompanied by criticisms that standards are

ditions, which have developed in part from the extension of education upwards, and in part from the intensive competition due to the economic situation following the World War, all point to the conclusion that the central problem in the administration of national systems of education is, to quote Sir Graham Balfour's statement again, "To enable the right pupils to receive the right education from the right teachers, at a cost within the means of the State, under conditions which will enable the pupils best to profit by their training." This statement may be paralleled by the terse summary of the problem confronting France in her efforts to give reality to the plans for the *école unique*. In discussing this problem M. Anatole de Monzie, former Minister of National Education, stated it could only be solved by *Gratuité, Sélection, et Rationalisation*—a system of free education implies selection and distribution of pupils; to this summary there should have been added *Orientation* or guidance to make it complete. In the same connection a Committee appointed to inquire into the *école unique* recommended the creation in the Ministry of National Education of a Permanent Commission for Selection and Guidance (*Commission Permanente de Sélection et Orientation*).

The crucial problem of the day centers, therefore, not so much upon the question of improving the machinery of examinations. As far as the individual pupil is concerned, he will be no better off educationally no matter how reliable may be the techniques by which he just manages to pass or fail. The ultimate question is the validity of the type of education from which he is capable of profiting. To improve the techniques for securing reliability of scoring while retaining a rigid curriculum may be a step in advance, but it would still mean concentration on the machinery of examinations alone. The task, however, is to read and interpret the results of examinations in such a way as to discover what is best for the individual's education and through the individual for the society in which he lives. The validity of an education will accordingly have to be defined in terms of the needs both of the individual and of society.¹

being lowered. Italy has attempted to meet the situation through the Gentile reform which added to the variety of schools at the secondary level but still bases admissions empirically on the traditional examinations. More recently the Soviet Republics of Russia, which entered blithely on a scheme of educational opportunities without demanding any qualifications other than political, have adopted a rigid system of annual examinations from the early elementary school years on. Japan has for several years been struggling with the problem of overcrowding in the universities, which has resulted in large numbers of students and graduates "harboring dangerous thoughts."

¹ The extent to which individual difficulties could be met by better methods of instruction, smaller classes, personal attention, and a breakdown of the lockstep time requirements under which all pupils are compelled to complete a given course in the same amount of time is a question which requires further investigation. The uniform time requirement has been abandoned in the University of Chicago experiment.

The attack on the problem of examinations in the United States began with the application of statistical methods to prove their unreliability in the traditional form. They began to be replaced by tests which would be more objective, more accurate, and more reliable. In time it was recognized that, even though this end was achieved, the result was still in terms of "pass" or "fail." It has been realized only within the last two decades that tests cannot be ends in themselves but can best serve their purpose as instruments for diagnosis, remedial work, and, finally, guidance. As instruments for educational guidance they have found their place and have won greater confidence. But even as instruments for educational guidance it is now generally admitted that tests cannot tell the whole story, that their predictive value is limited, and that if used only once their value is not much greater than that of the single examination of the traditional type.

This realization has led to an emphasis on the continuous use of tests over a period of years as a guide to the intellectual growth and development of a pupil. Beyond this has come the recognition that for adequate guidance the results of tests must be supplemented by more detailed information about the pupil's personality as a whole. Hence the adoption of the cumulative record card, continuous, comprehensive, detailed, which should provide a record of teachers' judgments, school marks, results of comparable tests which reveal rate and directions of growth, health, character, interests, and social conditions.

The adoption of the new type examination was not accepted without criticism. In the zeal to create a science of education, comparable with the physical sciences, too much was claimed for the new instruments by the early enthusiasts. Confronted by a certain inertia and opposition on the part of those who were not willing to give up the traditional examinations without a struggle, the enthusiasts, trained in the techniques, began to draw up tests which the subject matter specialists could not accept. Only gradually has there been reached in the construction of tests a rapprochement which has resulted in cooperation between the technical expert and the subject matter specialist. Nor is it yet claimed by those who are qualified to speak with authority that the existing tests have reached perfection. They are still subject to the criticism that they test "mere knowledge," but, as has been pointed out earlier, while the acquisition of knowledge is not the whole of education, an education without the acquisition of enduring and usable knowledge is inconceivable. Further, the criticism that the tests test only knowledge is denied and there is ample evidence to prove that even the recall

of knowledge involves reasoning, and that the tests are not merely tests of knowledge but also of the higher intellectual processes.

Again, those who are interested in education as a whole as well as in the techniques of measurement have at no time claimed that the new type tests are a complete substitute for other forms of examinations, training, and instruction. To prove that the marking of an essay examination is likely to be subjective and unreliable is not to deny the value of training in the writing of essays, even though that may be one of the unfortunate but unsupported results of the test movement. It is at this point that further developments in those countries which have been investigating their systems of examinations will prove of value to American educators. For the real test of an education in the European school systems is not only mastery of knowledge but ability to organize and communicate it, and, so long as this continues to be the test, it is not likely to be surrendered readily. Nor should it be.

On the other hand, if the results of the investigations into examinations carry conviction, these countries will be confronted with the task of combining the best in the traditional education with the best that can be found in the way of more accurate, reliable, and objective substitutes for the traditional examination. It will be to this solution that American educators will have to devote their interest and attention. For the European educators, however, the contribution which the United States can make lies not so much in the substitutes which have been developed for the traditional examination. For them the most important contribution lies in the conclusion that these substitutes, adequately supplemented with other information, constitute instruments that can be employed for the educational guidance of increasing numbers of young people whose education is beginning to create a social problem without parallel and without precedent. And yet not without parallel or precedent in theory! Plato had already given some attention to the problem and sought to provide the right education for the right individual in the just state on the basis of such psychological differentiation as he had at his command. But for the present study there is no more fitting and appropriate statement than John Locke's analysis of the problem with which the educational world must struggle in the twentieth century:

God has stamp'd certain Characters upon Mens Minds, which like their Shapes, may perhaps be a little mended; but can hardly be totally alter'd and transform'd into the contrary.

He therefore, that is about Children, should well study their Natures and Aptitudes, and see, by often trials, what turn they easily take, and

what becomes them; observe what their Native Stock is, how it may be improved, and what it is fit for: He should consider, what they want; whether they be capable of having it wrought into them by industry, and incorporated there by Practice; and whether it be worth while to endeavour it. For in many cases, all that we can do, or should aim at, is to make the best of what Nature has given; to prevent the Vices and Faults to which such a Constitution is most inclined, and give it all the Advantages it is capable of. Every one's Natural Genius should be carried as far as it could, but to Attempt the putting another upon him, will be but Labour in vain: and what is so Plaister'd on, will at best sit but untowardly, and have always hanging to it the Ungracefulness of Constraint and Affectation.¹

Three hundred years later, in 1892, President Charles W. Eliot presented the same problem from the modern point of view:

To discriminate between pupils of different capacity, to select the competent for suitable instruction, and to advance each pupil with appropriate rapidity, will ultimately become, I believe, the most important functions of the public-school administrator—those functions in which he or she will be most serviceable to families and to the state.²

But by no one has the problem been more succinctly stated than by Sir Graham Balfour in his definition of the function of the administration of education, a definition which has practically served as a text for this volume:

To enable the right pupils to receive the right education from the right teachers at a cost within the means of the State under conditions which will enable the pupils best to profit by their training.

This is ultimately the problem with which examinations and the distribution of education must concern itself.

¹ John Locke, *Some Thoughts Concerning Education*.

² C. W. Eliot, *Educational Reform*, p. 261 (New York, 1898).

APPENDIX

APPENDIX

THE INTERNATIONAL EXAMINATIONS INQUIRY

This summary of the International Examinations Inquiry was read by the author before the Fourth Educational Conference of the Committees on Personnel Methods and on Educational Testing of the American Council on Education, the Commission on the Relation of School and College of the Progressive Education Association, the Cooperative Test Service, and the Educational Records Bureau, and is reprinted from The Educational Record, January, 1936, Supplement No. 9

IT is a generally accepted fact in comparative education that educational systems, theories, and practices are determined by the ethos of the people whom they are designed to serve—their cultural, social, and political traditions, values, and aspirations. And yet, there is no country in the world today in which there are not discernible evidences of widespread unrest in education, and the crucial problem in this unrest is how to discover and provide the right education for the right individual under the right teacher. Whatever the backgrounds may be, whether one starts from the Right or Left politically, whether there is accepted a rigid definition of the purposes of education and the meaning of culture or an absence of a standardized currency of educational values, the central and most pressing problem appears to be the same,—the provision and distribution of education consistent with the needs of society and adapted to the aptitudes of those who are to be educated. The problem is not new, Plato struggled with it and within the limits of the psychology of his day attempted to provide a solution for his ideal state. The national systems of education which began to be created early in the nineteenth century solved the problem simply by providing two types of education—one for the masses and another for a select minority. This general form, based on social stratification, was continued and, when the progress of the nineteenth century demanded variation and differentiation, there developed an elementary branch with its ramifications upwards and the secondary branch leading to the universities and professional preparation. Here and there slight and tenuous bridges were thrown across the two systems, over which the ablest could find their way from the lower to the higher branch.

The rapid changes in the social, political, and economic realm which characterized the early years of the twentieth century and whose tempo was quickened by the Great War directed attention to education more intensively than ever before in its history. It began to be generally recognized that an elementary education in itself provides an inadequate preparation for life in modern society, the widespread acceptance for a short time after the War of the democratic ideal brought with it demands for increased equality of educational opportunities, the intensification of international competition brought with it the recognition that more education must be provided, and in those countries which abandoned the democratic ideal the recognition of the value of extended education in the interests both of national welfare and the rapidly increasing body of unemployable youth helped to save the

demand for a more generous distribution of education. The result is seen in the slow but inevitable abandonment of the dual system of education—one for the masses and another for the select group—and the adoption of a common foundation on the basis of which differentiated types of post-primary education are to be built.

The American tradition has proceeded on different lines. The nineteenth century opened with the acceptance, at first as an ideal and later as a reality, of the principle of equality of educational opportunity which resulted in the adoption of a unitary, articulated system from the lowest grade of school to the university. It was not, however, until the opening of the present century that the full implications of this ideal were realized. Dissatisfaction with what was described as the domination of college entrance requirements opened the way to greater flexibility in the programs of the high school, a flexibility which was still further increased by studies of elimination, mortality, and individual differences. The expansion of the high school offerings has proceeded apace and has only served to bring new problems of educational adaptation and adjustment in its train. Apparently something more was needed than the mere provision of an extensive *à la carte* educational menu, leaving the pupil to choose as he pleases.

Thus allowing for the differences in backgrounds and cultural traditions and starting from two extreme points of view the European countries and the United States seem to be converging toward the same problem—the differentiation and distribution of education. The European countries are on the eve of providing differentiated types of schools for differentiated abilities; the United States is confronted with the task of adjusting education to the needs and capacities of individual pupils but within the same institution. The European situation is, however, complicated by two factors which are not as potent and influential in the United States. The first of these is the jealous respect for a strong tradition of culture and liberal education, which unfortunately no longer exists in the United States. Secondly, the European problem is complicated by what may be called the concept of education for status—the social and economic valuation placed upon an education which leads to certain preferred occupations and professions and which under present conditions has brought with it overcrowding in the universities and the liberal professions. The problem thus resolves itself in all countries into the necessity of defining and providing the right education for the right individual under the right teacher, and this means, although it is not seen as clearly in some countries as in others, the necessity of devising methods whereby differences in aptitudes may be discovered.

In so far as attempts have been made to discover such differences the sole reliance in the past and with few exceptions in the present has been upon examinations of the traditional type. Indeed, it might almost be said that if Socrates were to return to earth he might gain the impression that his beautiful statement "The life which is unexamined is not worth living" had been accepted too literally and somewhat extravagantly, at least for that part of mankind which is under educa-

tional tutelage. (Lest we in the United States be too complacent on this subject, it may be well to point out that the late Dr. Carl Becker, former Minister of Education in Prussia, stated at the first Conference on Examinations that the number of examinations in Germany, reputedly the land of examinations, "is considerably less than in the United States of America, where students are being perpetually examined.") . . .

Under the sponsorship of the Carnegie Corporation, Dr. Monroe, as director of the International Institute of Teachers College, Columbia University, was appointed to organize and administer what came to be known as the International Examinations Inquiry.

The first Conference on Examinations was held at Eastbourne, England, in May, 1931, and was attended by delegates invited because of their known interest in the subject from England, France, Germany, Scotland, Switzerland, and the United States. There were twenty-eight delegates in all, a number which made round-table discussion possible. The absence of prepared papers, except some giving general descriptions of the educational systems of the countries represented, conduced to an informality and spontaneity of discussion which ranged over the whole field of education, general and vocational. To the student who is interested in understanding the educational values of the nations represented I can suggest no better source material than can be found in the eloquent statements reported in the *Conference on Examinations* (Bureau of Publications, Teachers College, Columbia University, 1931) The technician may not learn much from this report about the scientific conduct of examinations, but to the American technician the report can be recommended if only to obtain an understanding and appreciation of the matrix in which the problem of examinations is closely embedded.

The purposes of examinations were succinctly analyzed in the address with which Dr. Monroe opened the Conference. Examinations may be used (1) as a means of instruction, (2) as a form of educational administration; (3) for purposes of admission to various occupations, professions, and government service; and (4) as a means of social control. Three aspects of the problem appeared to dominate the discussion. The first centered round the question of social justice in the distribution of education for the development of individual talent, the second was concerned with the question of the selection and maintenance of an *élite*, the third was devoted to discussions of the meaning of culture, which ranged from the insistence of the French group on the importance of safeguarding *culture générale* against attack to the recognition that civilization is changing and our concept of culture must change with it. From the French point of view it was admitted that some experimentation with examinations is desirable but that ultimately the undisputed ideal must prevail. From the opposed point of view the pithy statement was made by Professor C. Delisle Burns that "One of the worst troubles in the whole examination system is that it has been devised by professors, and the best thing that professors can think of is themselves; they therefore test candidates by what are tests of competence for pro-

fessors, but not for bankers and other persons." Changes are going on all around us, "and we are still clinging to the old idea that we must produce persons and test them in the old way."

If little was contributed, except by Professors Thorndike and Judd, on the technical aspects of examinations, it may be said by way of general summary that the discussions at the First Conference helped to make clear that the problem of examinations can only be attacked after the purposes of education, its meaning and organization, have been clarified. If one may hazard an opinion in the light of the development of examinations and their substitutes in this country it is along these lines that the method of approach at the First Conference has a distinct contribution to make to American educators. It is significant that at no time in the seven sessions of the Conference was a voice raised in favor of dispensing altogether with examinations of some kind. The crux of the problem was stated by one of the delegates from Scotland, where greater progress has been made than elsewhere, largely through the efforts of the Scottish Council for Research in Education, in the use of scientific techniques of measurement. Education authorities in Scotland were already convinced that the existing system of examinations is unsatisfactory, but the question which they put is "What have you by way of substitute?" a question to which a small group in Germany was already beginning to provide an answer in the development of examinations that are both just to the pupil and have predictive value.

The value of the Conference is best described in the words of one of the distinguished delegates, Sir Michael Sadler, who said:

Never in a long life have I been to an educational conference more interesting or more impressive than this. We have met here from several countries and speak different languages and yet understand one another so perfectly that we might come from different states in one union. We are facing the same problem. We cherish the same hopes. We are all thinking of the future of humanity. The differences of language have simply added to the richness and interest of the discussion.

The result of the Conference was the unanimous recognition that here is a problem which is at once technical and, even more important, social in its ramifications, that it cannot be attacked from a unitary point of view, that public opinion on the subject has not reached the same stage in all countries, and that each country must develop its own resources and methods for investigating the problem. To this end national committees were appointed in each of the countries represented at the Conference to study the problem in ways most appropriate to their own situation. The further continuance of the Inquiry was made possible, as was the first Conference, by the generosity and interest of the Carnegie Corporation and the Carnegie Foundation.

National committees were accordingly organized in England, France, Germany, and Scotland. In Switzerland Professor Bovet undertook the preparation of a study

of the long history of the examination of army conscripts. The Committees in England and France had greater obstacles to overcome than the Scottish and German Committees. The Scottish Committee continued its studies along lines already initiated by the Research Council. Germany was already confronted by a very practical and very urgent situation, secondary schools and universities were overcrowded, the university authorities professed to see a lowering of standards in the post-War students, and the professions were also overcrowded; and the German Republic was in any case more ready for experimentation in education than either France or England. In the latter country Sir Philip Hartog, the Director of the Inquiry and long a critic of the system of examinations, in seeking the coöperation of one of the universities in the investigation was met with the statement, "We think that here we know everything there is to be known about examinations", in the former M. Desclos, the Director of the French Inquiry, attributes the failure of more than 2,800 out of 3,000 teachers and administrators to reply to a questionnaire on examinations to the fact that they have no doubt as to the impeccability of the findings of the examiners, who can be trusted to continue their work to the satisfaction of all concerned.

The task of the English and French Committees was accordingly defined by the situation, they had to prove in the most accurate and convincing manner possible the unreliability of marking in the existing practice of examinations in order to shake the prevailing complacency and before further investigations looking to reform could be undertaken. It was already evident at an Interim Conference of committee directors, held in London in June, 1933, that the members of at least one of the committees had themselves already received a greater shock to their complacency than they had expected in 1931, a fact which only strengthened their resolution to probe further. It may be claimed that the English and French Committees were only repeating studies and techniques on an aspect of examinations already closed and settled in the United States from the scientific point of view nearly a generation ago. This objection only brings into the foreground the fact that the educational problems can only be solved in the light of local cultural and other backgrounds.

The final reports of the investigations which had been conducted over the previous four years were presented at the Third Conference on Examinations held at Folkestone, England, in June of this year. The same countries were represented as at the First Conference with the addition of observers invited from Finland, Norway, Sweden, and Australia. The Third Conference differed from the First only in the fact that prepared reports were presented, as at the First Conference the advantages of intimate discussions and questions from a variety of backgrounds and familiarity with the issues involved again helped to clarify the problems raised.

The English Committee had devoted its investigations to testing concurrences of a number of independent examiners and of independent boards of examiners in marking the same set of papers, the purpose of the second study was to test the claim that individual idiosyncrasies of examiners are ironed out when the marking

is done by a board. As compared with earlier experiments conducted in this country the papers used in the investigations were actual papers written in actual examinations. These papers covered a wide range of examinations, both qualifying and selective, and included papers written in the eleven plus or special place examination, a college scholarship examination, university honours examinations in history and mathematics, and secondary school certificate examinations in Latin, French, chemistry, and history. An investigation was also conducted on the method of interview on the model of the Civil Service Examination. Examiners were selected from actual examining panels, were paid, and were given all the time that they needed. Every known variation of the examining technique was used in order to insure accuracy—the same system of marking was employed as in the original examination; the same examiners marked the same papers twice at an interval of a year; two groups of examiners were employed on the same papers, general standards and details of marking were discussed in two groups of examiners under the direction of chief examiners, trial markings were submitted to criticism, an English essay was marked once by general impression and then by details; and numerical and literal marks were given.

The French Committee proceeded on its investigations along much the same lines but concentrated more specifically upon the most crucial of French examinations—the *baccalauréat*, which comes at the close of secondary education. One hundred papers in six of the subjects of the examination—French essay, Latin translation, English, mathematics, physics, and the dissertation in philosophy—were each submitted to marking by five examiners. The papers were marked in the customary way and each examiner handed in reports on methods of marking employed, qualities and defects looked for, their relative importance, and the proportion of the total marks allotted to the several qualities discovered. In another study a French essay was marked by 76 examiners.

The detailed results of these investigations cannot be announced in advance of their publication by the committees concerned. Anyone familiar with similar studies will guess correctly that there was discovered a great degree of unreliability in the marking, that the fate of candidates in examinations may differ with the different examiners, that on the whole candidates are exposed to the unfairness of chance. A number of side issues emerged in the discussion of the results—What is the pedagogical function of examinations? Do examinations exercise any effect upon the work of the schools? Is the marking of an examination an art or a scientific process? What would be the effect of new methods of examining upon the traditional values in education? Can examinations of any kind succeed in measuring all the qualities of a good education? How can justice be done to pupils of different abilities either by existing or by new methods of testing? It is obvious from the discussions at both the First and Third Conferences that the subject of the Inquiry transcends the mere question of new devices and techniques of examination and that the technician must work in close coöperation with the educational theorist and statesman,

and that even then the question, as seen by the French and English educators, still remains whether what is most valuable in education can be measured. It is for this reason that the English Committee became interested in studying the method of the interview and the French are planning to undertake a study of their oral examinations, as well as to strengthen the results so far obtained by investigations into the examinations for the *certificat d'études primaires* and for national scholarships. The English Committee began its investigations with some anticipation of the results that would be reached and with the intention of expressing these results in such a form that they would carry conviction to other educators and to the public. The French have emerged from the Inquiry conducted up to the present "more modest and less dogmatic" about the impeccability of their examinations.

The Scottish Committee addressed itself to a somewhat larger question within the scope of the Inquiry. Before the Inquiry was initiated the Scottish Council for Research in Education had already undertaken two inquiries into teachers' estimates and the time devoted to examinations which were completed and published in 1932. Two new studies were undertaken, the one prospective, the other retrospective. The prospective inquiry consisted of a mental survey of Scottish children, representing a complete age-group of 87,498 children or practically all the children in Scottish schools born in 1921. The test used was Professor Godfrey Thomson's Group Intelligence Test. Of the total group 1,000 children, constituting a representative sample of the whole age-group, were further tested by means of the Terman revision of the Binet-Simon scale. The results of this study were published in 1933 under the title *The Intelligence of Scottish Children: A National Survey of an Age-Group*. The Examinations Inquiry Committee decided to keep in touch with the thousand children who had been tested individually and to follow them up through their school careers and, if possible, beyond. Such a follow-up study, somewhat similar to the Pennsylvania Study, will, it is hoped, result in discovering how far the promise at the age of eleven was fulfilled in their later careers.

The retrospective inquiry undertook to investigate the prognostic or predictive value of the examination for entrance to the university. The group investigated consisted of 472 students, men and women, who had entered one of the universities in 1928. The data available consisted of teachers' and principals' estimates, the marks obtained in the Leaving Certificate Examination, an external examination conducted under the auspices of the Scottish Education Department, marks given by professors in university class examinations, and marks obtained in the final examinations for a degree, either ordinary or honours. Only a few of the results can be given here: the principals' estimates were, on the whole, higher for those students who obtained honours than for those who did not, and were higher for students who obtained first class honours than for the second class honours, and so on. The composite marks in the Leaving Certificate Examination were in the same way predictive as between students who took honours or ordinary degrees. Finally, there was found a lower correlation between teachers' estimates and Leaving Cer-

tificate marks in certain subjects than between class and degree marks obtained in the university. The results of this study have also been published under the title, *The Prognostic Value of University Entrance Examinations in Scotland*. Evidence of the interest in the two Scottish studies is indicated in the cooperation of every education authority and private school, the central authority, and the university. It is hoped through these channels to influence public opinion on the one hand and to see that no pupil of ability fails to receive the education appropriate for him.

Of the German Committee only one representative attended to present the final reports of a series of studies. As in Scotland a group of educators had already begun to investigate the problems of examinations before the International Inquiry was instituted. The study of these problems had been stimulated in the German Republic by the need of discovering some adequate methods of selection for admission to the secondary schools, on the one hand, and by the tendency of German youth to crowd into the universities in the intense struggle for economic survival. The problems were as much sociological as they were technical. On the technical side the investigations at the stage of admission to the secondary schools were devoted to studies of the prognostic values of various methods of selection—teachers' estimates, school marks, entrance examinations, psychological tests. The general results of a series of investigations conducted in Berlin by Bobertag and Hylla and in Bremen by Valentiner produced a number of conclusions which had already been anticipated in similar studies in the United States. Teachers' estimates were in general inconsistent, but might be improved somewhat by special courses to teachers, the prognostic value both of school marks and of marks in entrance examinations was slight and unreliable; intelligence tests had a higher prognostic value than any other measure, and a combination of the three methods produced the best results. One investigation attempted to discover the qualities necessary for success in secondary education and resulted in the conclusion that intelligence was a more important factor than social or emotional qualities. Another study on the marking of an essay, while it showed agreement among several examiners on the very best and the very worst, revealed great variation between these points. All the studies have been published under the title *Schülerauslese* by O. Bobertag.

The sociological aspects were inherent in many of these studies and were the objects of one in particular. They constituted the main purpose of an investigation conducted by Dr. Robert Ulich and Dr. Erich Wohlfahrt, who undertook to show the relations between the social origin of students, the types of secondary schools selected, their professional aims, and their success in the courses undertaken. Social origin, as expressed in differences of cultural background, affected both selection of schools and careers and educational success. The general conclusion of the authors was, however, that there is no justification for the educational segregation of youth according to their social origin, and that from the national point of view methods must be devised whereby the ablest students of all classes, some stimulated by their cultural background and others by the will to rise socially, be selected to

preserve the "hereditary culture carefully nursed and maintained for generations", otherwise a nation is in danger of cultivating mediocrity. The study further revealed, both from the sociological and educational points of view, that the different types of secondary school courses differ in quality and standards and that the same is true of different departments in the universities. An analysis of examination results at the university level showed that the stricter selection was made in departments conducting fewer but more comprehensive examinations than in departments with frequent examinations during the course. School records were found to be unreliable as prognostic measures of success in the university, satisfactory results were attained by a combination of school records and selection by expert committees which took into account not only intellectual achievements as shown in school records but supplemented these by judgments on the whole personality of the candidates. An analysis of the investigations in this field undertaken under the auspices of the *Studienstiftung des Deutschen Volkes*, an organization established after the War to provide help for highly gifted students. The studies of the *Studienstiftung* centered around the problems of overcrowding of students in the universities and the liberal professions, the consequent competition and maladjustments, and the urgent need of adequate measures of selection and distribution of students. That these problems are just as pressing under the Nazi as under the Republican régime is indicated by the law of April, 1933, and subsequent decrees, and by an investigation undertaken by the Minister of Education in Saxony, published under the title *Geist und Torheit auf Primanerbanken*.

The American educator may be tempted to ask what contribution can the studies and research carried on in foreign countries make to the solution of his problems. On the technical side he may claim that he knows all that there is to know about the unreliability of the traditional examinations and their making, he has in the main passed the stage of inquiry reached by the foreign educators and has already made considerable progress in the development and elaboration of more or less reliable and valid substitutes, he has little to learn perhaps about the provision and organization of education because the ideal of equality of educational opportunities has for long been the basis of the American system of education. And yet one may answer that our own problems become clearer in the light of the discovery of similar problems elsewhere. The countries discussed in this article, and many other countries, are discovering that the crucial problem is not primarily one of selection but of the distribution of education according to aptitudes and abilities. Despite the great amount of work done on the technical side in the United States, one may hazard the statement that it has not been guided by a thorough recognition of purposes. The work in the United States has been influenced by a desire to discover those studies by which an individual is best capable of profiting. The European countries are just as much interested in this phase of the problem, but as they dig deeper into it, they realize that over and above individual differences of aptitudes and needs and urges the most important task still remains the definition and

preservation of educational values. The one solid contribution which comes out of the European inquiries is that the technical, psychological, statistical studies of examinations are as much concerned with the needs of society as of the individual, and that these studies must remain meaningless unless they are guided by a philosophy of educational values and purposes.

The statement was made earlier that the problems here discussed are not new. They were recognized by Plato; in the seventeenth century in the infancy of modern science the possibilities of applying scientific methods to the study of the human being were already envisioned. Bishop Thomas Sprat, the first historian of the Royal Society of London for the Improving of Natural Knowledge, after describing what science might ultimately contribute to human welfare, wrote in 1667:

In men, may be consider'd the *Faculties*, and operations of their *Souls*, The constitution of their *Bodies*, and the *works of their Hands*. Of these, the *first* they omit: both because the knowledge and direction of them have been before undertaken, by some *Arts*, on which they have no mind to intrench, as the *Politicks*, *Morality*, and *Oratory*: and also because the *Reason*, the *Understanding*, the *Tempers*, the *Will*, the *Passions* of Men, are so hard to be reduc'd to any certain observation of the *senses*; and afford so much room to the *observers* to falsifie or counterfeit: that if such discourses should be once entertain'd; they would be in danger of falling into *talking*, insted of *working*, which they carefully avoid. Such subjects therefore as these, they have hitherto kept out. But yet, when they shall have made more progress, in *material* things, they will be in a condition, of pronouncing more boldly on them too. For, though Man's *Soul*, and *Body* are not onely one *natural Engine* (as some have thought) of whose motions of all sorts, there may be as certain an accompt given, as of those of a Watch or Clock: yet by long studying of the *Spirits*, of the *Bloud*, of the *Nourishment*, of the parts, of the *Diseases*, of the *Advantages*, of the accidents which belong to *humane bodies* (all which will come within their Province) there, without question, be very near gheses made, even at the more *exalted*, and *immediate* Actions of the *Soul*, and that too, without destroying its *Spiritual* and *Immortal* Being.

And Socrates in his last hour faced his fate without flinching because "Above all, I shall be able to continue my search into true and false knowledge; as in this world, so also in that, I shall find out who is wise, and who pretends to be wise and is not." (Jowett's translation.) The question still remains whether, in the words of Bishop Sprat, in an inquiry into the immediate actions of *Soul*, this can be done "without destroying its *Spiritual* and *Immortal* Being," or, in modern terms, whether the search for the right education for the right individual under the right teacher by old or new type examinations can be conducted without destroying educational values.

It is at this point that the work of the Educational Records Bureau and the Cooperative Test Service derives its significance and meaning. This work is not an attempt to dilute standards or values, but, to the degree that measuring instruments have been perfected, it represents a genuine and sincere effort to discover

the real aptitudes of the student as an individual and to carry him along the lines on which he is best fitted to advance. It does not seek to discard educational values, but aims, so long as these are generally accepted, to develop them in their essential richness to the extent that a student can profit from them. In so far as these efforts succeed, they will tend to reduce, and perhaps eventually to eliminate, much of the tragic waste which still exists in education.

The acknowledged and proved existence of individual differences is frankly accepted and a method is provided by which they can be met more effectively. To accept as a principle that *non omnia possumus omnes* does not simplify the task of educators, but it places the onus where it belongs—on the recognition that one of the most serious responsibilities confronting educators today is that of guidance, which is possible only with the fullest knowledge over a long period of the potentialities of each individual to be educated. To the best of my knowledge neither the Educational Records Bureau nor the Coöperative Test Service nor other organizations engaged in similar work expect to answer all the problems of education; in the minds of enlightened leaders there is no suggestion that testing is a substitute for good teaching or that new type tests provide a substitute for good teaching or that new type tests provide a substitute for the best in the old type, such as the essay. Nor does the work of such organizations seek to compromise with educational values by positing some miracle of “thinking” without the wherewithal with which to think, nor does it seek to set up a fallacious distinction between the “how” and the “what” of education, but recognizes that in true education these cannot be separated. No one would claim that this work provides the ultimate solution of all problems in education, but it is a step toward discovering an answer to many of the problems raised at the Conferences of the International Examinations Inquiry.

INDEX

INDEX

ACCREDITING, 104 f.

- Advisory Committee on College Testing, 121
- Aikin, W. M., 54, 133
- Alternative courses, 12, 14 f.
- American and European education, 5 ff.
- American Council on Education, 121, 126, 129, 142, Psychological Examination, 96, 129, 132
- American Historical Association, 45
- American Philological Association, 45
- Amherst College, 36
- Analogies tests, 87
- Anderson, H. R., 129
- Appel, F. S., 102
- Aptitude tests, 53 f.
- Arithmetic tests, 75
- Army Alpha Test, 95
- Army conscripts, examination of, 169
- Association for Education in Citizenship, 13
- Association of College Presidents in Pennsylvania, 134
- Association of Colleges and Secondary Schools of the Middle States and Maryland, 36
- Association tests, 87

Baccalauréat, 89; report on, 155 f.

- Bagley, William C., 76
- Balfour, Sir Graham, 20, 159, 162
- Barnard, H., 24, 27
- Becker, Carl, 167
- Beers, F. S., 83
- Berechtigungen*, 11
- Binet, 74
- Bobertag, O., 172
- Boston, examinations in, 22, 24 f.
- Boston University, 36
- Bovet, P., 168
- Briggs, T. H., 10, 19
- Brigham, Carl C., 55
- Broome, E. C., 23, 37
- Brown University, 36
- Brunschwig, Léon, 16
- Buckingham, B. R., 71
- Burk, F., 71
- Burnet, Mr., 157
- Burns, C. Delisle, 17, 167
- Butler, Nicholas Murray, 36, 44

CALDWELL, O. W., 24 ff.

- Carnegie Corporation of New York, vii f., xi, 167 f.
- Carnegie Foundation for the Advancement of Teaching, vii, 134 f., 168

Carstater, E. A., 101

- Cattell, J. McKeen, 59, 73 f.
- Central Committee on Personnel Methods, 121
- Certificat d'études primaires*, 171
- Chapin, Angie C., 58
- Chicago, University of, 103 f.
- Classification tests, 87
- Coffman, L. D., 103, 117
- College Aptitude Rating, 117
- College Cumulative Record Form, 145 f.
- College Entrance Examination Board, 7, 36, 43 ff., 59, 104, 130, 156 f.
- College Psychological Examination, 129
- College Sophomore Testing Program, 121 f.
- College success, prediction of, 53 f., 95 f., 106, 116
- Columbia College, 95
- Colvin, Stephen S., 76 f.
- Comin, R., 65
- Commission on the Relation of School and College, 133
- Commission Permanente de Selection et Orientation*, 159
- Committee of Review, 45
- Committee of Revision, 45, 47, 53
- Committee of Ten, National Education Association, 37
- Committee on Educational Testing, 121 f.
- Committee on Examination Ratings, 46
- Committee on Uniform College Entrance Requirements, 37, 45
- Commonwealth Fund, 129
- Complehon tests, 87
- Consultative Committee, England, 16
- Contemporary Civilization, course in, 99
- Cooperative Test Service, 122 ff., 126 ff., 131, 174 f.; list of tests, 122 ff.
- Corrective work, tests and, 131
- Courtis, S. A., 24 ff.
- Cowley, W. H., 72
- Coxe, W. W., 39
- Crawford, Margaret Elnor, case of, 146 ff.
- Credits, 12, 134 ff.
- Cross-out tests, 87
- Culture générale*, 167
- Culture tests, general, 137
- Culture, tradition of, 8, 11, 152 f.
- Cumulative record cards, 20, 135 f., 140 ff.

DARTMOUTH College, 36

- Dearborn, W. F., 63
- Denmark, 158
- Desclos, A., 104

Determinism in tests, 76
 Diagnostic tests, 131 ff.
 Differentiation in education, 166
 Discipline, formal, 9
 Dispersion, 33 f.
 Distribution of education, 159
 Draper, Andrew S., 38
 Duffus, R. L., 12

EASTBOURNE, Conference on Examinations, xi, 3, 167
Ecole unique, 14
 Edgeworth, F. Y., 34 f.
 Education and examinations, 151 ff.
 Education, common problems of, 13 f.
 Educational guidance, 133, 140 ff., 143 ff.
 Educational Records Bureau, 124, 130 ff., 174 f.
Einheitsschule, 14
 Eliot, Charles W., 30 ff., 37 f., 162
Elite, selection of, 167
 Elliott, E. C., 61 f
 England, examination inquiry in, 169 f.
 England, secondary education in, 11 f., 14
 English, examination experiment in, 61 f
 Enrollments, increasing, and tests, 93 f.
 Equality of opportunity, 70, 124 f., 144
 Essay examination, 85, 101 f.; marking of, 66 f.
 Euclid, A. C., 102 f.
 European and American education, 5 ff.
 European education, aims of, 173
 Examination boards, 36 ff
 Examinations, College Entrance Board, 43 ff.
 Examinations, Conference on, xi, 3, 167 ff.
 Examinations, criticism of, 57 ff., external, 104 ff., 156; issues in, 170, purpose of, 70; reliability of, 158; reports on, 154, scientific attack on, 69 ff.; social setting and, 3 ff.; and education, 5, 151 ff.; traditional, 22 ff.; value of, 57 ff., 66; written, 37
 Examinations, Regents', 37 ff.

FAILURES, 89; in education, 70 ff., 89; problems of, 15
 Farrand, Livingston, 74
 Fees in secondary schools, 15
 Finkelstein, I. E., 63
 Fiske, John, 89
 Fiske, Thomas S., 48
 Flexner, Abraham, 151, 153
 Fluctuations in examination results, 49 ff.
 Folkestone, Conference on Examinations, 169
 Formal discipline, 9
 France, examination inquiry in, 170 f.
 Frutchey, F. P., 99

GALTON, Sir Francis, 73 f.
 Gates, A. I., 92
 General College, University of Minnesota, 100 f., 117 ff.
 General Education Board, 126
 Gentile, 159
 Geometry, examination experiment in, 62
 Germany, examination inquiry in, 172
 Germany, secondary education in, 11
 Goddard, H. H., 74
 Godley, A. T., 28
 Grading, reliability of, 61 ff.
 Grizzell, E. D., 24
 Guidance, 20, 159 f.; educational, 133, 140 ff.

HADLEY, Arthur T., 58
 Haggerty, M. E., 100
 Hancock, W. K., 18
Handwriting Scale, 75
 Harris, J. Rendell, 33 f.
 Hartnacke, W., 16, 19, 72, 158
 Hattog, Sir Philip, 4, 19, 169
 Harvard College entrance examination, 23
 Hawkes, H. E., 91
 Henmon, V. C., 113
 High school, meaning of, 10
 Hildreth, G., 93
 History, examination experiment in, 62
 Hitler, 15
 Holt, F. O., 113
 Hopkins, L. T., 51
 Hylla, E., 172

INDEPENDENT schools, 130
 Individual differences, 9
 Information, tests of, 83 f.
 Intelligence quotients, 20
 Intelligence tests, 54, 135 f.
 International comparative studies, 133
 International Examinations Inquiry, 19 f., 165 ff.
 International Institute, Teachers College, Columbia University, vii, ix, xi, 167
 Interview, method of, 170 f.
 Iowa Every-Pupil Achievement Testing Program, 109 f
 Iowa Silent Reading Test, 132
 Iowa statewide examinations, 109 f.
 Iowa, University of, 109 f.
 Italy, 159

JAPAN, 159
 Jessup, Walter A., xi
 Johnson, P. O., 102
 Johnston, J. B., 117, 119 f.
 Jones, E. S., 121
 Judd, C. H., 168

KANDEL, I. L., vii f., 3, 5, 8 f., 11, 16, 37, 69, 104, 144, 158
 Kelly, F. J., 64
 Kent Education Committee, 145
 Keppel, Frederick P., ix, xi
 Kilpatrick, W. H., 81
 Knight, Edgar W., 144
 Knowledge, mere, and tests, 138
 Knox, W. W., 37, 57
 Koos, L. V., 10
 Kotschnig, W., 153
 Krey, A. C., 102

LANG, A. R., 87
 Lapie, P., 7
 League of Nations' Mission of Educational Experts, 12
 Learned, William S., vii, xii, 5, 83, 134 f., 138, 140
 Leaving Certificate Examination, 171 f
 Leicester, examinations in, 24
 Lindquist, E. F., 110, 129
 Locke, John, 161 f.
 Lowell, A. Lawrence, 82
 Lowell, examinations in, 24

MACGREGOR, G., 134
 MacLean, M. S., 84
 McCall, W. A., 87
 McConn, Max, 91, 128, 138 f., 144
 Maine, University of, 36
 Mann, C. R., 91
 Mann, Horace, 24 f., 28
 Marks, variability of, 62 f.
 Massachusetts, examinations in, 22
 Matching tests, 87
 Methods, statistical, in education, 160 f
 Meyer, Max, 63
 Minnesota statewide examinations, 115 ff
 Minnesota, University of, 92, 100 ff.
 Modern Language Association, 45
 Monroe, Paul, viii f., xi, 167
 Monroe, W. S., 66, 87
 Monzie, Anatole de, 159
 Morrison, S. E., 23
 Morrison, Henry C., 72
 Mount Holyoke College, 36
 Mulhern, J., 24, 28
 Multiple-Choice tests, 87
 Multiple-Response tests, 87

NATIONAL Bureau of Selection and Guidance, 89
 National Education Association, committees of, 37, 45

National Research Council, 129, 145
 National Society for the Study of Education, 74
 Nationwide testing programs, 120 ff
 Nelson-Denny Reading Test, 132
 New Education Fellowship, International Commission, 145
 New England Association of Colleges and Preparatory Schools, 36
 New England College Entrance Certificate Board, 35 f.
 New England Colleges, Conference of, 36
 New Type tests, 39 ff., 79 ff., 135 f., 140 f., 160 f.
 New York State Regents' Examinations, 37 ff., 95, 104
 Nichols, E. H., 58
 Normal curve, 69
 Norms, use of, 131
 North Central Association, 61
 Northern Universities Joint Matriculation Board, 7, 11, 157
 Norway, 158

OBJECTIVE tests, 39 ff., 79 ff., 135 f., 140 f., 160 f.
 Odell, C. W., 65, 87
 Ohio College Association Intelligence Test, 106
 Ohio statewide examinations, 105 f.
 Orata, P., 9
 Otis tests, 132
 Oxford and Cambridge Schools Examination Board, 157

PALMER, F., Jr., 72
 Paterson, D. C., 88
 Patterson, B. F., 28
 Pennsylvania College Achievement Test, 122
 Pennsylvania State Department of Public Instruction, 134
 Pennsylvania Study, vii f., 133 ff., 146
 Philadelphia Academy, 23
 Pintner-Cunningham Primary Mental Tests, 132
 Pintner, R., 74 f.
 Plan A, College Entrance Examination Board, 53
 Plan B, College Entrance Examination Board, 53
 Plato, 161, 163, 174
 Points, 12
 Portsmouth, examinations in, 24
 Powers, S. R., 133
 Predictability, criteria of, 95 ff., 106, 116
 Prediction of college success, 53 f., 95 f., 106, 116
 Pritchett, Henry S., vii
 Private schools, 130
 Progressive Achievement Tests, 132
 Progressive Education Association, Commission on the Relation of School and College, 133

Prospective inquiry and examinations, 171
Public and education, 9
Pupil's Record Card, Kent, 145
Pupils, selection of, 172

QUALITY in education, 134 f.
Quantitative measures in education, 134 ff.
Questions, preparation of examination, 46 f.
Quincy, J., 23

RAUBENHEIMER, A. S., 98
Reasoning and tests, 83 f.
Recall tests, 87
Recognition tests, 87
Record card, cumulative, 135 f.
Reeves, W. F., 96
Regents' Examinations, New York State, 37 ff., 95, 104
Reifepfung, 158
Reliability of marking examinations, 51, 61 ff., 158
Remedial work, tests and, 131
Retrospective inquiry and examinations, 171 f.
Rhodes, E. C., 4, 19
Rice, J. M., 75
Richardson, H. D., 10 f.
Ruch, G. M., 64, 87
Russell, C., 87
Russell, J. D., 96, 104
Russia, 159

SADLER, Sir Michael, 168
Salem, examinations in, 24
Sanctions, 11
Sandiford, Peter, 64, 78
Savage, Howard J., xii
Scholarships, national, 171
Scholastic aptitude tests, 54, 56 f.
Scientific attack on examinations, 69 ff.
Scientific movement in education, 57 f., 73 f.
Scotland, examination inquiry in, 169, 171 f.
Scottish Council for Research in Education, 168 f.
Secondary and Higher Education in Pennsylvania, Study of, 133 ff.
Secondary education, meaning of, 10
Secondary School and College Cumulative Record Forms, 143 f.
Secondary Schools Examinations Council, 154
Segel, D., 105, 107, 131
Selection, 152; methods of, 19 f.
Similarities tests, 87
Simon, 74
Smallwood, M. L., 23
Smith College, 36

Smith, John Morton, Jr., case of, 148 f
Social justice and education, 167
Social setting and examinations, Ch. I
Social studies, 13
Sociates, 106, 171
Spender, J. A., 8
Sprat, Thomas, 174
Stalnaker, J. M., 66 f., 104
Stalnaker, R. C., 66 f.
Standardizing agencies, 104
Standards in education, 8, 114
Starch, D., 61 ff., 65
Statewide testing programs, 104 ff.
Stone, *Arithmetic Tests*, 75
Strang, R., 93, 95 ff., 150
Streeter, Superintendent, 27
Students, scholarship, 173
Success, college, prediction of, 53 f., 95 f., 106, 116
Suzzallo, H., 22
Sweden, 158
Switzerland, examination inquiry in, 168 f.
Symonds, P. M., 87, 93

TEACHERS, secondary school, 12
Testing programs, nationwide, 120 ff., statewide, 104 ff.
Tests, 20; and instruction, 86, 127, 133; and more knowledge, 138; and reasoning, 83 f.; aptitude, 53 f.; Army, 75 f.; constricting influence of, 127 f., 133; new type, 79 ff.; objective, 79 ff., 160; production of, 128; use of, 131
Thomson, Godfrey, 171
Thorndike College Entrance Examination, 95 f.
Thorndike, E. L., 10, 59 ff., 74 f., 78 ff., 168
Thurstone, L. L., 104, 129
Thurstone, Thelma G., 129
Tildsley, John L., 11, 144
Toops, H. A., 105
Touton, F. C., 98
Tradition, American, 166
Tradition, the Great, 152
Traxler, A. E., 68, 82
Tufts College, 36
Tyler, R. W., 82, 99

ULICH, R., 172
United States, aim of education, 173; characteristics of education in, 163
United States and Europe, educational systems of, 5 ff.
Units, 12

VALENTINE, C. W., 19, 158
Validity of education, 152

Variability, 124, 137 f.; of marks, 62 f
Vermont, University of, 36
Vittorino da Feltre, 152 f.

WEAVER, R. B., 68, 82
Weidemann, C. C., 85
Wellesley College, 38
Wesley, E. R., 102
Wesleyan University, 36
Wesson, R. R., 102
West Riding of Yorkshire, 12, 15
Whitcraft, L. H., 57
White, E. E., 28 f.
Whitman, A. D., 53

Williams College, 36
Williamson, E. G., 117, 119
Wisconsin statewide examinations, 112 ff.
Wohlfahrt, E., 16, 19, 72, 158, 172
Wood, Ben D., xii, 39 ff., 64 f., 71, 77, 82, 87, 90,
95, 99 f., 127, 135, 145
Woodward, W. H., 153
Wyhe, A. T., 74

YALE College, examinations in, 23
Yerkes, R. M., 76
Yoakum, C. S., 76

ZOOK, G. F., 120