

The Century Education Series

TRADITIONAL EXAMINATIONS AND NEW-TYPE TESTS

BY

C. W. ODELL, PH.D.

ASSISTANT DIRECTOR,
BUREAU OF EDUCATIONAL RESEARCH,
UNIVERSITY OF ILLINOIS

THE CENTURY CO.
New York & London

PREFACE

In the preparation of this book the needs of teachers actually in service have been foremost in the writer's mind, though it has been his intention also to make the book serviceable to prospective teachers and others interested in the field covered. Its purpose is to present both the traditional and the new-type examination in rather detailed fashion, to point out the merits and the limitations of each, and to suggest how each may be so constructed and used as to yield the greatest returns. It would perhaps have been desirable to include standardized tests as well as examinations made by the teacher, but anything approaching a complete treatment of both is hardly practicable within the limits of a single volume. It seemed more practicable because of the much smaller amount of space required, and also more essential, to include a discussion of school marks. Accordingly two chapters have been devoted to this topic. The remainder of the book deals rather directly with the two general types of examinations already referred to, the traditional and the new-type.

In view of the very considerable amount of attention devoted to testing in the educational literature of the past few years, it is possible that readers will ask what justification there is for another treatment of the subject. Standardized tests have received the lion's share of this recent attention and the new-type tests most of the rest. The treatment of the traditional examination has been mostly incidental and has consisted largely

PREFACE

of destructive rather than constructive criticism. Moreover, most of the mass of material dealing with new-type tests is in the form of periodical articles, pamphlets, monographs, single chapters in books, and so forth and is, therefore, decidedly fragmentary. The chief merit claimed for this book is that it is based upon a comprehensive review of the material just referred to, and therefore, it is hoped, has gathered together the best of this material into a single unified treatment, rather than that it is to any considerable extent an original contribution.

C. W. ODELL

EDITOR'S INTRODUCTION

Almost all who are engaged in educational work, either as students or teachers, are vitally interested in the subject of marks and examinations. While a not inconsiderable number of instructors are prone to say that they have no confidence in the results of tests of any kind and wish that they could be eliminated, there is doubtless little probability that there will be any radical change in the means by which recognition is given for the successful completion of work in education. It is also fully as probable that for a long time to come instructors will feel compelled, in order to assure themselves from time to time of the character of the work which is being given in their courses, to offer some kind of so-called test designed to furnish this information.

This situation makes the publication of a book on marks and the new examinations especially timely. For a number of years a large amount of experimentation has been in progress looking to the formulation of new types of examinations which may be assumed to be more objective in character than the traditional examinations and which may lessen, if not eliminate, some of the evils which are incident to the use of the older type of test. A large number of educational institutions, ranging from the elementary schools to the professional schools, are making a more or less extensive use of the various types of the so-called new examinations. The author of this volume has made an unusually

thorough study of the whole subject and presents a definitely organized statement and explanation of all of the more commonly known types of new examinations and many of the modified types which have not received as wide publicity and have not been used to nearly as great an extent.

The subject is of such vital importance for all classes of school work that it seems highly desirable that a popular presentation should be available to all types of educators. While there is very little reason to believe that the educational world is going to abandon the traditional examinations, because there are certain educational values obvious to all in their use, there is little doubt that many of the newer types of examinations as illustrated in this volume will be used with increasing frequency. All those engaged in education should be well informed concerning these variants, their weaknesses and their advantages.

In the belief that this volume will prove of real value in extending widely the intelligent use of objective tests and the more critical and desirable selection of questions for discussion along traditional lines, this volume is submitted.

C. E. CHADSEY

CONTENTS

CHAPTER	PAGE
I THE PAST AND PRESENT STATUS OF EXAMINATIONS	3
1 INTRODUCTION. A BRIEF HISTORY OF EXAMINATIONS AND OF THE CRITICISMS THEREOF . .	3
2 ADVERSE CRITICISMS OF EXAMINATIONS . . .	9
3 THE DEFENSE OF EXAMINATIONS AGAINST THE CRITICISMS JUST GIVEN	11
4 ADVANTAGES OF STANDARDIZED TESTS OVER EXAMINATIONS PREPARED BY THE TEACHER . . .	19
5 ADVANTAGES OF EXAMINATIONS PREPARED BY THE TEACHER OVER STANDARDIZED TESTS . . .	27
6 SUMMARY	30
II WHAT ARE GOOD EXAMINATIONS? . .	32
1 THE PURPOSES EXAMINATIONS SHOULD SERVE .	32
2 THE QUALITIES OF GOOD EXAMINATIONS . . .	40
3 SUMMARY	58
III HOW TO MAKE AND GIVE EXAMINATIONS	59
1 THE PREPARATION OF GOOD EXAMINATIONS . .	59
2 THE ADMINISTRATION OF EXAMINATIONS . . .	64
3 SUMMARY	80
IV SCORING PUPILS' RESPONSES	81
1 THE WEIGHTING OF EXERCISES	81
2 SCORING EXAMINATION AND TEST PAPERS . .	86
3 CHANGING SCORES INTO MARKS	97
4 SUMMARY	107

V	THE MARKING SYSTEM AND ITS MEAN- ING	109
1	SHOULD MARKS BE USED AT ALL?	109
2	UPON WHAT SHOULD MARKS BE BASED?	113
3	WHAT MARKS SHOULD BE EMPLOYED?	133
4	SUMMARY	139
VI	THE DISTRIBUTION OF MARKS	141
1	SHOULD MARKS FOLLOW THE NORMAL OR ANY OTHER FIXED FREQUENCY DISTRIBUTION?	141
2	SUGGESTED PRACTICES CONCERNING THE USE OF A STANDARD DISTRIBUTION OF MARKS	149
3	ASSIGNING MARKS TO PUPILS IN SELECTED OR NON-AVERAGE GROUPS	159
4	ADJUSTING THE MARKS OF TEACHERS WHO DO NOT CONFORM TO THE STANDARD	164
5	SUMMARY	173
VII	MERITS AND LIMITATIONS OF TRADI- TIONAL AND NEW-TYPE EXAMINA- TIONS	175
1	MERITS AND ADVANTAGES OF TRADITIONAL, AND LIMITATIONS AND DISADVANTAGES OF NEW- TYPE, EXAMINATIONS	175
2	MERITS AND ADVANTAGES OF NEW-TYPE, AND LIMITATIONS AND DISADVANTAGES OF TRADI- TIONAL, EXAMINATIONS	183
3	SUMMARY	203
VIII	EXAMPLES OF TRADITIONAL OR ESSAY EXAMINATIONS	205
1	TYPES OF MENTAL ACTIVITY TO BE TESTED	205
2	EXAMPLES OF ESSAY QUESTIONS CALLING FOR THE TWENTY TYPES OF MENTAL ACTIVITY NAMED ABOVE	207
3	EXAMPLES OF GOOD DISCUSSION QUESTIONS IN LITERATURE	210
4	A COMPLETION ESSAY EXAMINATION	211

5	TRADITIONAL EXAMINATION QUESTIONS SELECTED FROM THOSE ACTUALLY USED BY PUBLIC- SCHOOL TEACHERS	221
6	TRADITIONAL EXAMINATION EXERCISES BASED UPON QUOTATIONS	233
7	SUMMARY	236
IX THE CONSTRUCTION AND USE OF NEW- TYPE EXAMINATIONS		
1	CONSTRUCTING NEW-TYPE EXAMINATIONS	237
2	SCORING NEW-TYPE EXAMINATIONS AND HAN- DLING THE RESULTS	246
3	THE SELECTION OF THE MOST APPROPRIATE TYPES FOR CLASS-ROOM USE	248
4	SUMMARY	253
X SINGLE-ANSWER OR RECALL TESTS		
1	GENERAL DISCUSSION	255
2	ORDINARY SINGLE-ANSWER TESTS	258
3	SINGLE-ANSWER TESTS EACH CONTAINING ONLY ONE EXERCISE	261
4	SO-CALLED "ASSOCIATION" SINGLE-ANSWER TESTS	263
5	DEFINITION OR DESCRIPTION SINGLE-ANSWER TESTS	267
6	SINGLE-EXAMPLE TESTS	270
7	PLURAL OR MULTIPLE-EXAMPLE TESTS	272
8	COMPOUND SINGLE-ANSWER TESTS	277
9	SUMMARY	279
XI MULTIPLE-ANSWER TESTS		
1	GENERAL DISCUSSION	281
2	ORDINARY MULTIPLE-ANSWER TESTS	289
3	PLURAL MULTIPLE-ANSWER TESTS	310
4	COMPOUND MULTIPLE-ANSWER TESTS	316
5	MULTIPLE-REASON TESTS	322
6	MULTIPLE-DESCRIPTION TESTS	329
7	SUMMARY	332

XII	ALTERNATIVE TESTS	334
1	GENERAL DISCUSSION	334
2	TRUE-FALSE EXERCISES	347
3	YES-NO QUESTIONS	350
4	OTHER VARIETIES OF TESTS HAVING ONLY TWO POSSIBLE ANSWERS	351
5	ALTERNATIVE TESTS WHICH PROVIDE A THIRD POSSIBLE ANSWER	354
6	SUMMARY	356
XIII	COMPLETION TESTS	358
1	GENERAL DISCUSSION	358
2	SIMPLE COMPLETION TESTS	362
3	COMPLETION TESTS WITH SUGGESTED ANSWERS	366
4	OTHER VARIETIES OF COMPLETION TESTS	372
5	SUMMARY	374
XIV	MATCHING TESTS	376
1	GENERAL DISCUSSION	376
2	ORDINARY MATCHING TESTS	378
3	COMPOUND MATCHING TESTS	384
4	SUMMARY	387
XV	INCORRECT-STATEMENTS TESTS	388
1	GENERAL DISCUSSION	388
2	EXAMPLES OF INCORRECT-STATEMENTS TESTS	390
3	SUMMARY	394
XVI	MISCELLANEOUS TYPES OF THE NEW EXAMINATION	395
1	GENERAL DISCUSSION	395
2	IDENTIFICATION TESTS	395
3	DISTINGUISHING TESTS	403
4	CONTINUITY OR REARRANGEMENT TESTS	405
5	VERIFICATION OR JUDGMENT TESTS	412
6	ANALOGIES TESTS	416
7	SUMMARY	420

XVII	OBJECTIVE TESTS IN INSTITUTIONS OF HIGHER LEARNING	422
1	GENERAL DISCUSSION	422
2	OBJECTIVE TESTS IN LIBERAL ARTS COURSES . .	424
3	OBJECTIVE TESTS IN PROFESSIONAL COURSES. .	430
4	SUMMARY	435
	BIBLIOGRAPHY	439
	INDEX	463

**TRADITIONAL EXAMINATIONS AND
NEW-TYPE TESTS**

TRADITIONAL EXAMINATIONS AND NEW-TYPE TESTS

CHAPTER I

THE PAST AND PRESENT STATUS OF EXAMINATIONS

1. **Introduction.** A brief history of examinations and of the criticisms thereof. Examinations are not new. They are in no sense a modern fad or a newfangled procedure recently introduced into school practice. On the contrary, they have been employed as a regular and integral part of school work for hundreds, even thousands, of years. A merely casual study of the educational practices of the past is sufficient to yield the information that the use of examinations of some sort or other has been the rule rather than the exception. In most countries, and in most of the schools in these countries, tests ¹ of achievement have had their place. Probably the best known, and certainly one of the most elaborate, of the ancient systems of examinations was that maintained by the Chinese; but the Greeks, the Romans, and many other peoples employed them as well. In fact, definite tests of what has been learned have had their place in practically every scheme of education of which we have any record. Many of them

¹ For the sake of variety, the word *test* will be used synonymously and interchangeably with *examination*.

were not written or even verbal, being rather tests of ability to apply and to do, but none the less they were examinations designed to test mastery of what was supposed to have been learned. When we come to medieval education, we find numerous accounts not only of public disputations and other testing procedures different from those commonly employed at present, but also of others very similar to the examinations with which all readers are undoubtedly familiar. In some cases these exercises, whatever their nature, were taken seriously by both teachers and pupils, whereas in others they were mere formalities, but the name and practice were general. Not only were they employed in connection with school work, but also, notably in China, they found a place outside the school-room for purposes which are now mostly included under the term *civil service*. At present, examinations are probably more prevalent than ever before, being used by practically all teachers in rating pupils, by government departments in choosing officials, by numerous business concerns in selecting and promoting employees, and by various other agencies.

Not only are examinations an institution of long standing, but also adverse as well as constructive criticism of them is nothing new. To go no further back than the eighteenth century, one finds an Oxford student of 1766 quoted to the effect (2,² p. 16) that his examination for the degree was an absolute farce, consisting of one very easy question in Hebrew and one in history. In this country Horace Mann, writing in 1845 (10, p. 37), stated certain points of superiority of what he called the "new" method of examining. Presumably,

² The numbers in parentheses refer to references listed in the Bibliography beginning on p. 439.

therefore, he had a more or less unfavorable opinion of the examination methods in general use in his time. Mann especially praised the use of uniform written questions instead of oral ones, which were usually different for different pupils. Apparently, however, no such volume of criticism of both the function and form of examinations as has appeared within the last few years ever arose during any previous period, and seemingly few persons at any one time in the past believed that anything serious was wrong with current practices regarding examinations or that any major change was needed. Therefore the form of examinations and the methods of using them underwent comparatively little change during a long period and only recently have modifications of any importance occurred or even been strongly urged. Although the last half-century or less has witnessed unusual advances in providing better buildings and material equipment, constructing more adequate curricula, and improving the technique of instruction, supervision, and administration, examinations were little affected during most of this period, but continued to exist in essentially unchanged form and to be employed in practically unchanged circumstances. It is only within the last few years that they have been subjected to any considerable amount of critical evaluation and discussion and that the twofold question as to their proper function and best form has been raised.

Undoubtedly the chief cause contributing to raise the question just mentioned was the publication of the results from a number of investigations which showed, or appeared to show, great unreliability³ and varia-

³ *Reliability* may be briefly defined as accuracy in measuring whatever is measured. For a fuller discussion, see p. 41.

bility of the marks given examination papers by teachers. Prominent in making the studies referred to were Johnson (40), Starch and Elliott (81, 82, 83), Kelly (41), and Dearborn (17). Their work, and also

Marks	Number	that of others along this line, is too
90-94.....	2	well known and has produced too
85-89.....	7	similar results to justify detailed ac-
80-84.....	11	counts of the various studies here. A
75-79.....	26	single striking example, which is also
70-74.....	13	probably the best-known one, will
65-69.....	18	suffice. Starch and Elliott (82) sub-
60-64.....	17	mitted an examination paper in ge-
55-59.....	8	ometry to a number of teachers of
50-54.....	5	mathematics, and were able to secure
25-49.....	7	the marks given the paper by 114

teachers, each marking independ-
ently. The accompanying distribution of marks is re-
ported. The actual marks ranged from 28 to 92 per cent,
75 being the passing mark. One or two examples even
more extreme than this have come to the writer's atten-
tion, and many others showing nearly as great varia-
bility have been mentioned in educational literature.
Altogether a large mass of data has been accumulated,
published, and generally accepted as valid, which offers
strong evidence that if pupils' answers to examination
questions of the type almost exclusively used until very
recently are marked by the same person a second time
after he had forgotten his first marks, or by two or
more individuals working independently, very great
differences in the marks assigned will frequently result.

It is only fair to say, however, that although the
men named above and others have presented the many
data referred to, their conclusions have not been ac-
cepted by all persons interested as representative of

ordinary conditions in the schools. In general, those who do not admit the validity of the conclusions drawn maintain that the marks collected and used were not assigned under ordinary school circumstances, but rather under conditions which tended to produce much greater variability than is the case in general school practice. Not only has the evidence for great variability been criticized, but also some data have been presented which indicate a decidedly high degree of reliability or agreement among teachers. Bolton (5), for example, reports on the scoring of twenty-four arithmetic papers by twenty-two Seattle teachers. The papers were scored by the ordinary percentile system and the variations from the average mark for each paper computed. The average variation was approximately 5 per cent. About one-sixth of the variations were not greater than 1 per cent, and one-third more, or one-half in all, not greater than 3 per cent. Further data presented by the same writer show that in the marking of single arithmetic questions almost a third of the teachers agreed with the average rating exactly, and that more than another third differed by not over 1 per cent. Furthermore, Bolton analyzes some of the data which Starch used as evidence of great variability, and as a result claims that they actually show decided uniformity in marking.

Even though it is admitted that the evidence which shows great unreliability and variability among teachers' marks is not conclusive, the fact remains that it has been generally enough accepted as valid to create a wide-spread doubt as to the reliability of marks as commonly given. The question of how to improve the situation by constructing examinations, training teachers, or both, so that much closer if not