

The Improvement of the Written Examination

BY

G. M. RUCH

ASSOCIATE PROFESSOR OF EDUCATION AND PSYCHOLOGY
STATE UNIVERSITY OF IOWA



SCOTT, FORESMAN AND COMPANY
CHICAGO ATLANTA NEW YORK

Copyright, 1924

BY

SCOTT, FORESMAN AND COMPANY

2511.4

GENERAL PREFACE

Educational tests are needed for a number of reasons, among the most important of which are the following:

1. To show the pupil the efficiency with which he has worked.
2. To show the teacher the efficiency with which she has taught.
3. To measure the value of a given textbook.
4. To measure the value of a given method of teaching.
5. To compare the efficiency of teaching in one school with that in another school.
6. To guarantee a thorough and rigorous attack, not only in study but also in teaching.
7. To get a general but useful index of the status of teaching in any given school system.

It is now clear that although standardized tests have performed and still perform a very useful service, they are not well adapted to many of the needs which have just been outlined. Few if any can be used to measure progress over a brief period. Many objectives of importance have no standardized tests by which they can be measured, nor are the standardized tests easily modified or adapted for the measurement of a changing course of study or of a course of study where differences in judgment exist as to what is important to teach. They are also influenced greatly not only by special practice upon the tests themselves, but also by exercises which closely resemble those contained in the tests. Therefore an improvement in the score on a given test cannot be taken as a safe indication of the progress really made in the set of abilities which has been tested. Some of the standard tests which are now on

the market have remained unchanged for many years so that the standards originally set are not now satisfactory.

This does not mean that standard tests have not served an important purpose or that they are not or will not continue to be of value. It merely shows the limitations of these tests. Of these various limitations the most important is the difficulty of adapting them to the measurement of accomplishment of specific objectives, and particularly those which are to be accomplished in a relatively brief unit of time, such as a day, a week, or a month. Often great injustice is done both teacher and pupil by presuming to measure an accomplishment by a standardized test which is not a valid measure of that accomplishment, and although it may be urged that standardized tests need not be so misused, the fact remains that such misuse is all too frequent. Certainly this is true in term testing. For example, efficiency in teaching spelling in a given term is commonly measured by a test selected without reference to the words taught during that term. Similarly, improvement in reading during the term is frequently measured by one or more standardized tests which correspond by no means perfectly to the specific habits, attitudes, and abilities which the pupil and teacher attempted to develop.

There can be no question but that standardized tests have had a very wide and beneficial influence in increasing the experimental study of school practice, as well as in the development of techniques and methods of objective testing. Through the interest which has been an outgrowth of the use of standardized tests, there has now developed a new type of test program which emphasizes the technique of testing rather than the test itself. These new and more perfect methods of testing fill a need which cannot be met by the use of standardized tests. In the first place, they are easily adapted to measure progress over any given unit of time no matter how small, and, second, they are adapted to the measurement of the accomplishment of any educational objective no matter how

small. They are therefore admirably suited for use in experimental work, either in the field of curriculum research or in the field of educational methods. They supplant not the standardized tests which are hardly suited for monthly, weekly, or daily testing, but rather the traditional essay test which, although intended for such use, is now known to be so unreliable as to be very unsatisfactory.

Dr. Ruch has performed a real service in presenting the evidence concerning the reliability of various forms of tests and in illustrating so clearly the different ways in which these tests may be prepared. There is no question, in the writer's judgment, but that this book marks out the next big field of improvement in testing the results of teaching.

ERNEST HORN,
Professor of Education,
State University of Iowa

AUTHOR'S PREFACE

This small volume has been prepared as the result of repeated requests from superintendents, principals, and teachers who are desirous of studying critically some of the newer proposals for the improvement of examination technique. In reality, it is an outgrowth of a syllabus prepared for use with a university class in objective examination methods.

It is hoped that the following pages include enough of the significant work which has been done thus far in developing objective methods to provide a bird's-eye view of the aims and progress of these efforts. Every attempt has been made to provide concrete examples and illustrations of possible examination techniques, especially in the fields of composition, literature, and reading which offer particular difficulty to the maker of examinations.

At times it has been necessary to introduce some elementary statistical concepts and practices. For the most part these have been placed in footnotes or in the final chapter in order that the reader may omit these more technical portions of the discussion if his interests are less specialized. It should also be said that certain statistical points have purposely been given a rather superficial treatment in order to avoid unnecessary technicalities.

The author's thanks are due Dr. Ernest Horn, who read the entire manuscript and made many helpful suggestions. Dr. M. F. Carpenter and Miss Helen Eddy read portions of the text and also contributed some of the examinations used for illustrations. Mr. George D. Stoddard prepared the graphs of Chapter III and checked some of the mathematical calculations. Specific acknowledgments to a number of the author's students have been made throughout the text.

G. M. R.

Iowa City, Dec. 1, 1924.

CONTENTS

CHAPTER I

	PAGE
FUNCTIONS OF WRITTEN EXAMINATIONS	1
Introduction.—Horace Mann as an Early Critic of Examinations.—Functions of Examinations.—Types of Examinations.—Examinations as Devices for Motivation.—Examinations and Language Expression.—Examinations as Measuring Instruments.—Criteria of a Good Examination.—SUMMARY.	

CHAPTER II

THE CRITERIA OF A GOOD EXAMINATION	13
I. VALIDITY. Definition of Validity.—Validation Illustrated in the Case of Reading.	
II. RELIABILITY. Definition.—Statistical Determination of Reliability.—Causes of Unreliability.	
III. OBJECTIVITY.	
IV. EASE OF ADMINISTRATION AND SCORING.	
V. STANDARDS.	
SUMMARY.	

CHAPTER III

SOURCES OF ERROR IN WRITTEN EXAMINATIONS	40
The Problem.—Previous Investigations.—Recent Investigations.—SUMMARY.	

CHAPTER IV

TYPES AND CONSTRUCTION OF THE NEWER OBJECTIVE EXAMINATIONS	65
Classification of Objective Examination Types.—EXAMPLES OF OBJECTIVE EXAMINATION TYPES.—THE CONSTRUCTION OF OBJECTIVE EXAMINATIONS: Steps in the Building of an Objective Examination.—The Validity of the Newer Objective Examination Methods.—Use of the Results of Objective Examinations in Grading Pupils.—SUMMARY.	

CHAPTER V

	PAGE
EXPERIMENTAL STUDIES OF SEVERAL TYPES OF OBJECTIVE EXAMINATIONS	106
The Relative Merits of the Various Types.—Experimental Studies of Objective Examination Methods.—Corrections for Chance.—Results of Similar Investigations.—Additional Pedagogical and Psychological Considerations.—SUMMARY.	

CHAPTER VI

STATISTICAL CONSIDERATIONS RELATED TO EXAMINATION TECHNIQUE	131
Introduction.— Problem 1: To Determine the Reliability Coefficient of a Written Examination.— Problem 2: To Determine the Amount of Error Present in an Individual Score or Mark on an Examination or Test.— Problem 3: To Determine What Constitutes a Satisfactory Reliability Coefficient for an Examination.—SUMMARY.	
BIBLIOGRAPHY.	149
APPENDIX.	
I. Ruch High-School Content Examination	154
II. Objective Examination in First-Year French	180
III. True-False Test in First-Year Latin	183
IV. Objective Examination in Manual Training	185
INDEX	191

CHAPTER I

FUNCTIONS OF WRITTEN EXAMINATIONS

Introduction. Examinations are almost or quite as old as formal education, and formal educational methods evolve but slowly. This is as it should be, since, after all, education is one of the conserving forces in society. But the fact that this virtue of slow evolution tends at all times to become a vice, with the consequent result that the best teaching practice tends to lag too far behind the best thought and theory of the moment, would probably be admitted by most educators. It is especially unfortunate that examination methods have persisted throughout the centuries with modifications that are almost microscopic in comparison with the much more conspicuous advances in such aspects of education as the conduct of the recitation, the course of studies, the psychology of the school subjects, and the administrative organization of our schools. It is not too much to say that only within the past quarter of a century have important reforms been introduced into our technique of measuring the efficiency of instruction.

The reasons for this inertia in the perfecting of methods of written examinations have been two-fold in character. In the first place, until very recently it has not been realized that there are any disputed issues involved in the usage of the traditional written examination. The inherent defects of the method have come into clear relief only with the publication of the studies of Daniel Starch, F. J. Kelly, F. W. Johnson, and others, on the unreliability of teachers' marks.* These criticisms have pointed the way to a consideration of many previously unsuspected issues. In the second place, the reforms resulting from these studies have been woefully

*See Chap. III, "Sources of Error in Written Examinations."

meager, because the conclusions were chiefly *negative* in character. We are met with the situation today that large numbers of teachers and school officers are justly suspicious of the worth of the typical written examination, without possessing adequate knowledge of the technique for eliminating these faults and dangers.

Horace Mann As An Early Critic of The Examination. The above statements are not quite literally true. As early even as the time of Horace Mann, careful thinkers had given attention to the matter of the validity of written examinations. For calling attention again to the views of Horace Mann on examination practices, we are greatly indebted to the authors of a most interesting book, *Then and Now in Education*.*

In Chapter IV, under the caption "Unanswerable Arguments from the Past," the authors write:

By no means the least interesting reading in connection with the 1845 report is the portion of the *Common School Journal* in which Horace Mann justifies the use of printed tests by the survey committee. He saw at once the meaning and value of objective measurement, and even now when the conflict for the scientific study of educational processes by means of standard tests has been won, his appraisal of the method of the first American survey is still most enlightening.

Briefly, Horace Mann states that the new method of examination has a decided superiority over any and all other methods, for the following reasons:

1. It is impartial.
2. It is just to the pupils.
3. It is more thorough than older forms of examination.
4. It prevents the "officious interference" of the teacher.
5. It "determines, beyond appeal or gainsaying, whether the pupils have been faithfully and competently taught."
6. It takes away "all possibility of favoritism."
7. It makes the information obtained available to all.
8. It enables all to appraise the ease or difficulty of the questions.

*Caldwell, Otis W. and Curtis, Stuart A. *Then and Now in Education, 1845-1923*. Yonkers-on-Hudson, N. Y.: The World Book Company, copyright. 1923.

To be sure, Horace Mann did not have in mind exactly the same type of objective examination as will be treated in this volume, but his arguments can be carried over bodily. Reference to the examinations used by the survey committee of 1845 will show that they were far from completely objective. But, after all, that was seventy-five years ago, and several important movements, like the rise of the standard educational test and the evolution of mental measurement, have intervened to shape methods of examining today. The interesting thing is that Horace Mann could, in 1845, write a valid set of arguments for objective examination practices that can be accepted today with but minor modifications and additions!

Returning to our main task, it may be stated that the function of this book is to suggest a methodology of examination construction which is immediately serviceable to the classroom teacher and supervisor. To accomplish this end, it will first be necessary to make a critical analysis of the aims and functions of the written examination as an educational instrument.

Functions of Examinations. The purposes of written and oral examinations have been held to be of three sorts:

1. *Motivation.* Educators seem to be agreed that pupils tend to accomplish more when confronted with the realization that a day of reckoning is surely at hand. The truth of this statement rests much more largely upon the consensus of opinion, it must be admitted, than upon any very convincing experimental evidence. Only within the past few years have the problems of motivation been held to be the subject for experimentation rather than for a priori beliefs. It is indeed fortunate that direct experiment is proving the fundamental truth of our assumptions about the role of motives in learning.

2. *Training in written expression.* The accomplishment of training in written expression is undoubtedly one of the important desiderata of education. Though the emphasis has been gradually shifting away from written expression and in the direction of oral expression, the change is merely one of a more

acceptable distribution of effort from the point of view of social utility, and implies no lack of faith in the values of written language in the curriculum. This purpose for the examination may therefore be admitted, with certain reservations to be brought forward later in the book.

3. *Measurement.* Measurement has generally been the most conscious aim of all. It has also been the one which has been the recipient of the brunt of recent attacks. Examinations and measurements are the instruments by which the results of education can be evaluated. Since educational theory has produced no alternative method of portraying the results of teaching, we are perfectly safe in admitting pupil-measurement as one of the legitimate functions of the written examination in so far as the examination can be shown to be a valid and reliable means to this end.

This inventory of the purposes served by written examinations does not pretend to be complete. Other purposes suggest themselves immediately, and many additional ones do possess obvious merit. The three aims described have been selected because of their relative importance and because of their direct bearing on the problems for discussion in this text. Before proceeding to a more comprehensive evaluation of these aims, it will clarify matters to suggest the types of examination methods which have been proposed and widely used up to the present time.

Types of Examinations. Historically, we can classify examination practices as of three kinds:

1. The traditional "essay-type" of written examination.
2. The standardized educational test or scale.
3. The newer objective examination.*

Each of these three types has its peculiar characteristics, advantages, and limitations. The superiority or inferiority of

*Objective examinations are to be defined as examinations where the grading does not involve the personal opinion of the grader. The pupil writes very little, but answers the questions by underlining or checking the correct answers. The best known types are the true-false, multiple-choice, matching, and completion exercises. Samples of all of these are given in Chapter IV.

any one of these methods can at best be stated only in terms of very specific ends to be attained. A blanket defense or condemnation of any one of them is quite impossible. Many partisan battles are yet to be waged over this question of relative merits, and, indeed, all such controversies are doomed to be miserably barren unless a clear definition of examination objectives can be marked out. Three such objectives have already been stated and accepted as valid for the purposes of the present treatment. Two of these will be dismissed with scant attention, for reasons which will now be stated.

Examinations as Devices for Motivation. With respect to the problem of the motivation value of examinations, there is no need for detailed treatment here. None of the proposed examination practices listed in the preceding paragraph is inimical to this crucial aim in teaching. Pupils learn partly by virtue of learning capacities which are native or acquired (apparently largely native according to our best evidence today), and partly, or very probably chiefly, in proportion to their interests, application, and volitional constitution—in other words, in accordance with their motivation. No one can say whether the situation of facing a series of questions like: “List the main causes of the American Revolution,” or “Why does the rainfall of the western slope of the Rockies exceed that of the eastern slope?” contains more dynamic or motivating properties than the situation of facing a series of exercises like:

Slavery was one of the important issues in the American Revolution. true false

Taxation without representation was one of the chief causes of the American Revolution. true false

The vegetation of the eastern slope of the Rocky Mountains is more luxuriant than that of the western slope. true false

In the absence of any experimental information on this point, we need say no more than to suggest that in general the motivation of learning is likely to be greatest in that situation where the pupil comes to realize that definitely organized knowledge

is expected rather than the throwing of a few scattered half-truths into the semblance of a paragraph structure. The bluffing, muddle-headed student finds himself much more at a disadvantage in the newer objective examination than he does in the older essay-type. If important differences are found in the motivating values of the older and the newer examinations, these differences are certain to favor the more objective methods, since the pupil rapidly comes to sense the exactness of the requirements of the newer examinations.

Examinations and Language Expression. Turning to the uses of examinations in the teaching of written language expression, it must be admitted that at first glance the argument seems against the standard test and the objective examination. There is no question about this in theory, at least. In actual school practice, however, the issue is not so clear. In the first place, the two newer types of examinations *do not claim to serve this purpose*. This admission is the less damaging because it is thoroughly honest and conscious. That it is a real limitation is quite as evident. No lack of faith in the desirability of training in direct, forceful, and logical expression of thought as a part of instruction and measurement alike is implied. It is nearer the truth to confess that in the newer types of examination an important function of the examination has been voluntarily sacrificed in the interest of other ends held to be equally important and more surely obtainable.

This suggests an opening for a critical examination of the usual written essay-type of examination to discover exactly what evidence supports its claim to value as language training. That written examinations *might* provide excellent opportunities for language training is certain; that they *do* accomplish much of such training is again a matter of faith, not proof. At least two arguments can be brought forward which weaken the claims of the traditional examination, the one practical, the other psychological.

The practical situation of the writing and correction of an

examination paper is usually about as follows: The pupil writes on a series of questions which, nine times out of ten, are far too long and comprehensive for completion in thirty or sixty minutes unless the resort is to a set of answers scribbled under high tension in the attempt to finish the job. The teacher reads the papers in much the same spirit. The task at best is uninteresting and monotonous after the first dozen papers have been completed. Final examinations in particular are likely to be corrected on the day that all the multitudinous tasks incident to the closing of a semester or year of school rush in on the tired teacher, who must finish the papers, make out the report cards, balance the register, pack her trunk, and catch the earliest train home. Where, then, does the opportunity for training in the correct use of English come in under the situation as it really is?

In the second place, it is psychologically unsound to believe that so complex an ability as language training can be built up as a by-product of the written examination. The known laws of learning are opposed to this belief. Pupils acquire such skills and abilities only under patient, specific instruction, and adequate practice. The learning must proceed by slow stages with constant helps over the difficulties. The attention must not be divided between the listing of facts and the effort at careful expression.

Pupils will probably continue to write examinations with the primary view of getting down the facts as quickly as possible. Teachers will probably continue to grade these papers with the primary purpose in mind of ascertaining whether a sufficient number of these facts have been set down by the pupil. Neither the pupil nor the teacher disbelieves in the virtue of a well-written paper from a literary standpoint, but neither is likely to exert great effort toward this end under the prevailing conditions of writing examinations. The situation is not made any better by the current practice in many schools of filing all final examinations in some convenient office without giving

opportunity for the pupils to profit by the results of teachers' comments, if indeed any are offered.

What is the solution? The linguistic functions of the examination must be saved. These are even more important than is generally realized. The problem will be clarified and certain suggestions will arise when we realize that our examination technique must provide for two different things: first, the discovery of the pupil's mastery of subject-matter, i. e., measurement of achievement; and second, training in thought and expression, written and oral. The traditional examination serves neither well. The standard test serves the first only. The newer objective examination will serve the first only.

The solution would seem to be that a new methodology must be created for language training and measurement apart from the more general examination for factual mastery. Just what this should be is not within the projected purpose of this text; nor (it must be confessed), is it within our present knowledge. Drills in written language usages, diagnostic exercises, specially derived quality and judgment scales in composition and literature, etc., are urgently needed and undoubtedly will ultimately be produced. The interests of progress will be better served by differentiating these two separate objectives and meeting both needs with examination instruments designed to these separate ends than by the continuation of our present muddled practices in which the supposed values are at least undemonstrated, if not largely illusory.

Examinations as Measuring Instruments. The way to the improvement of examinations from the point of view of the measurement of mastery of factual content is fortunately much more clear. The same promise holds for other types of educational objectives, like training in thought, judgment, and appreciation. These statements bring us to the consideration of the third of our specific aims of the examination, *viz.*, measurement.

Measurement has always been the prime consideration in

examinations. Although perhaps not a primary function of education, measurement does stand in the position of being the method by which the realization of primary objectives can be demonstrated. In this position it is rather unique among educational practices. Until recently no one has felt the necessity of challenging either the logic or the efficiency of the written examination. When we recall the storm of protest raised at the beginning of the present century in the meetings of the National Educational Association when one bold spirit, Dr. J. M. Rice, suggested that it would be quite defensible to state the efficiency and worth of the teacher in terms of the progress of the pupils, we wonder what conception of the purpose of examinations existed in the minds of educational workers.

There was surely a curious lack of logic in this situation of the examination providing an adequate and valid measure of the pupil, but at the same time being entirely inadequate and unfair as a measure of the teacher. To be sure this conservatism was directed at the newly derived instrument, the standard test, but indirectly it did involve the idea of measurement, since increased accuracy of measurement was the main claim of the standard test. Now that standardized examinations or tests have earned their place, it might appear that these issues are closed. The standard test has demonstrated time and again its peculiar worth, but it has not and probably will not supplant entirely the written examination. The latter has its legitimate place and its individual advantages. The relations of the two are mutually supplementary, not exclusive.

That the written examination is very imperfect is also to be concluded as the result of a long series of experimental studies, beginning with the investigations of the unreliability of teachers' marks to which reference has already been made at the beginning of the chapter. Kelly, Starch, Johnson, and others, have