

# EDUCATIONAL TESTS AND MEASUREMENTS

BY

WALTER SCOTT MONROE, PH.D.

PROFESSOR OF EDUCATION

DIRECTOR OF THE BUREAU OF EDUCATIONAL RESEARCH

UNIVERSITY OF ILLINOIS

ASSISTED BY

JAMES CLARENCE DEVOSS, A.M.

ASSOCIATE PROFESSOR OF PSYCHOLOGY AND PHILOSOPHY

KANSAS STATE NORMAL SCHOOL

AND

FREDERICK JAMES KELLY, PH.D.

DEAN OF THE SCHOOL OF EDUCATION

UNIVERSITY OF KANSAS

HOUGHTON MIFFLIN COMPANY

BOSTON NEW YORK CHICAGO SAN FRANCISCO

*The Riverside Press Cambridge*

**COPYRIGHT, 1917, BY W. S. MONROE, J. C. DEVOSS AND F. J. KELLY**

**ALL RIGHTS RESERVED**

**The Riverside Press**  
**CAMBRIDGE . MASSACHUSETTS**  
**PRINTED IN THE U.S.A.**

## EDITOR'S INTRODUCTION

UP to very recently our chief method for determining the efficiency of a school system was the method of personal opinion. When the work of a superintendent of schools was called in question, the schools were visited and personal opinions expressed as to their standing. In case of a disagreement among the visitors the efficiency became a matter of dispute, and the people of a community usually favored the opinion which most nearly coincided with their prejudices and preconceived ideas.

Relatively recently the method of comparison was introduced. By means of this method the school system under consideration is compared with other school systems of the same size and class, and with reference to a number of different items. After such a comparison has been made, it is possible to place the school system relatively. If the school system studied stands fourth out of twenty school systems compared in one item, thirteenth in another, and at the bottom of the list in three others, it is not difficult to determine its position. It is evident that this is a much better method than the one of personal opinion. Its chief defect, though, lies in that the school system studied is continually compared with the average or median of its size and class. In other words, the school system is continually measured as against mediocrity, when as a matter of fact the average or median school system may not represent a good school

system at all. Perhaps all of the school systems below the average or median should be classed as poor school systems, and even some of those above are not doing what a school system should do.

Still more recently, and wholly within the past decade, a still better method for the evaluation of the work which teachers and schools are doing has been evolved. This new method consists in the setting up, through the medium of a series of carefully devised "Standardized Tests," of standard measurements and units of accomplishments for the determination of the kind and the amount of work which a school or a school system is doing. This new movement is as yet almost in its infancy, but so important is it in terms of the future of school administration that it already bids fair to change, in the course of time, the whole character of this professional service.

The significance of these new standards of measurement for our educational service is indeed large. Their use means nothing less than the ultimate transformation of school work from guesswork to scientific accuracy; the elimination of favoritism and politics from the work; the ending forever of the day when a personal or a political enemy of a superintendent can secure his removal, without regard to the efficiency of the school system he has built up; the substitution of well-trained experts as superintendents of schools for the old successful practitioners; and the changing of school supervision from a temporary or a political job, for which little or no preparation need be made, to that of a highly skilled piece of social engineering.

This new method for the evaluation of the work which a

school system is doing is so important that any young man or woman of to-day who desires to prepare for school administration should by all means thoroughly familiarize himself or herself with the aims and methods of this new type of administrative procedure. The underlying purpose of the new movement has been the creation of such standardized scales for measuring school work, and for comparing the accomplishments of different schools and groups of school-children, as to give to both supervisors and teachers definite aims in the imparting of instruction. Instead of continuing to teach without definite measuring-sticks, and to assign tasks and trust to luck and the growth process in children for results, which is comparable to the old-time luck-and-chance farming, it has been attempted to evolve standards of measurement which will do for education what has been done for agriculture as a result of the application of scientific knowledge and scientific methods to farming.

Such an important new movement is of especial significance to the teacher in charge of a class, to the citizen interested in schools, and to the superintendent responsible for results.

To the teacher it cannot help but eventually mean not only concise and definite statements as to what she is expected to do in the different subjects of the course of study, but the reduction of instruction to those items which can be proved to be of importance in preparation for intelligent living and future usefulness in life. It will mean, too, an ultimate differentiation in training for the different types of children with which teachers now have to deal, and the specialization of work so as to enable teachers to

obtain more satisfactory individual results. To the citizen the movement means the erection of standards of accomplishment which are definite, and by means of which he can judge for himself as to the efficiency of the schools he helps to support. For the superintendent it means the changing of school supervision from guesswork to scientific accuracy, and the establishment of standards of work by which he may defend what he is doing.

Up to the present time nearly all of the work which has been done in the evolution and testing out of these new standardized tests has been work of a highly scientific and technical nature, most of the articles being written in a language which the layman can scarcely understand. Often no interpretation has been attempted of the results which have been obtained. The classroom teacher and the school principal have naturally not found these studies of much help to them in their work.

This work has been carried far enough, however, so that the time now seems ripe for a clear and simple statement as to the nature of the different tests which have been evolved, their use, their reliability, what are the best standard scores so far arrived at, and, in particular, how to diagnose the results and apply remedial instruction. This the three authors of the present volume in the series have attempted to give, and, to make their work of the largest possible usefulness to normal-school students, teachers, and principals of schools, they have cast the whole in language so simple and untechnical that the average grade teacher can read the book and understand it. In addition, to give still larger value to the book, they have added a

number of chapters, written in a similar simple and readable style, giving the essential elements needed in understanding simple statistical methods, the meaning of scores, the unreliability of school marks and their relation to standardized scores, and the use of the standardized tests in the work of school supervision.

No space has been taken up in merely reproducing the tests themselves, though samples, showing their nature, have been inserted. If it is desired to use the tests with a class, they will be needed in quantities, and they may then be obtained in quantities and for very small sums from the persons and at the places mentioned in the chapter bibliographies. These bibliographies also give the most important book or article describing in detail the construction and use of the tests, in case the worker desires to go further than this volume presents. Instead, the authors have used their space in explaining to teachers and school officers the nature of the tests, telling how to give and score them, what standings the pupils should attain in their use, and presenting a rather full description as to the significance of the results obtained and how to remedy the defective conditions which the use of the tests reveals. In consequence, the book should prove of much use not only to students in normal schools and colleges, but to teachers and principals in our public schools as well. The style and contents of the volume are such as also to adapt it to reading-circle study with teachers, or to the needs of the average citizen interested in knowing something as to the nature and uses of the Standardized Tests.

ELLWOOD P. CUBBERLEY.

## PREFACE

THIS book is designed primarily for teachers. It is based on two years' experience in giving a course on educational measurements to prospective teachers in a state normal school and on the experience received from directing a Bureau of Educational Measurements and Standards.

It is just twenty years since Rice startled the educators of this country by his proposal that the results of teaching spelling could be measured by a spelling test. His proposal was greeted with sarcasm and ridicule, but during the past two decades the opposition to the principle of educational measurements has almost entirely disappeared. To-day the widespread use of standardized tests and scales bears witness to the importance of this movement in American education. However, it is profitable to analyze our present interest in educational measurements. A thing may be interesting merely because it is new and spectacular. Scores are objective and are subject to graphical representation. A chart displayed attracts attention. Evidence is not wanting to show that a considerable number of teachers look upon educational measurements merely as an interesting topic for teachers' meetings or as a means of attracting attention in their community.

Standardized tests and scales are not "playthings." Neither are they teaching devices. They are instruments which furnish the teacher (1) with detailed and definite



aims, and (2) with a means for diagnosing the teaching situation which she faces. Unless the diagnosis is followed by remedial instruction the use of standardized tests and scales cannot be of much value. They become mere "play-things."

Our present tests are probably crude instruments, but the first railway locomotive was also crude. Even now standardized tests and scales are superior to ordinary examinations, but, more important, their use tends to engender in the teacher a type of thinking about her work which is very helpful. By using them she recognizes objective standards to be attained and not to be exceeded, the present achievements of her pupils, and that instruction must be suited to the needs of her pupils. When a teacher comes to think of her teaching problem in these terms, she is in a position to increase greatly her efficiency.

This book is addressed to teachers because they are charged with the instruction of pupils. The superintendent, principal, or student of education who is interested in the teacher's work also will find much of value in the book. Technical details of the derivation of tests are not given, but references are given so that one interested may pursue the matter. These were omitted because they are not essential to the use of the tests by teachers. For much the same reason the criticism of tests is made a secondary matter. The detailed criticism of tests and the derivation of improved ones must be left to the expert. The teacher needs to know only enough to enable her to choose wisely in selecting a test, and to prevent her from ascribing to the scores a significance which is not justified.

The newness of the field and the rapidity with which it is developing places limitations upon an attempt to write a text. It is recognized that probably before this volume is printed new tests will have been announced. However, the author believes that the point of view upon which the book is based is not merely temporary, and that, as new tests are available, the fundamental principles of the book may be applied to them.

It is obvious that in an endeavor such as this one must utilize the results obtained by many investigators. In fact it is hoped that this book may have the virtue of summarizing these results. The author is keenly aware of his obligation to all whose work is mentioned in the following pages. Special mention should be made of Professor DeVoss, who contributed the chapter on "Handwriting," and of Dean Kelly, who wrote the chapter on "Reading."

WALTER S. MONROE.

EMPORIA, KANSAS, *April 27, 1917*

# CONTENTS

## CHAPTER I. THE INACCURACY OF PRESENT SCHOOL MARKS . . . . . 1

School marks — The inaccuracy of teachers' marks — Carter's investigation — Kelly's investigation — Conclusions from these studies — Johnson's investigation — Marking examination papers — Distribution of marks — Error due to unequal value of questions — Another example of unequal values — Rate of doing work neglected — Wide range of topics included within an examination — Most valuable topics for education — Questions and topics for investigation.

## CHAPTER II. ARITHMETIC . . . . . 17

### I. THE PROBLEM OF MEASURING ARITHMETICAL ABILITIES.

Arithmetical abilities automatic or habits — Arithmetical abilities distinct — Separate types in handling integers — Each a specific habit — Why we need to use arithmetical tests.

### II. STANDARDIZED TESTS FOR MEASURING ARITHMETICAL ABILITIES.

1. The Courtis Standard Research Tests, Series B — Marking the papers.
2. The Cleveland-Survey Arithmetic Tests — Spiral nature of these tests — Nature of the tests.
3. The Woody Arithmetic Scales — Measurement by means of a scale — The addition scale.
4. Research Tests in Addition of Fractions.
5. The Stone Reasoning Test.
6. Other reasoning tests.

### III. STANDARD SCORES.

1. Courtis Standard Research Tests, Series B — Courtis Standard scores.
2. The Cleveland-Survey Tests — Cleveland and Grand Rapids scores.
3. The Woody Arithmetic Scales.
4. The Addition of Fractions Tests.
5. The Stone Reasoning Test.

The accuracy of individual scores — Gain or loss in repeating tests.

## CONTENTS

### IV. HOW TO HANDLE WHAT THE TESTS REVEAL.

Scientific management — Diagnosis of the teaching situation — Pupils' and class records charted — Meeting the situation; laws — Individual *vs.* class needs — Repeating the tests after an interval — Modifying the class drill — Use of practice tests — Questions and topics for investigation — Bibliography.

## CHAPTER III. READING . . . . . 66

### THE PROBLEM OF MEASUREMENT IN READING.

#### I. SILENT READING.

##### (a) Recognition of words.

1. The Thorndike Visual Vocabulary Scales — Use and standards.
2. The Haggerty Visual Vocabulary Tests.
3. Starch's English Vocabulary Tests.

##### (b) Tests for comprehension and speed.

1. The Thorndike Scale Alpha.
2. The Minnesota Scale Beta.
3. The Courtis English Tests.
4. Brown's Silent Reading Test — Value of the test in diagnosis.
5. Starch's Silent Reading Tests.
6. Gray's Silent Reading Tests.
7. The Kansas Silent Reading Tests — Score value for the Kansas tests — Standard median scores.
8. The Courtis Silent Reading Tests.

#### II. ORAL READING.

1. The Jones Visual Vocabulary Tests.
2. The Haggerty Visual Vocabulary Tests.
3. Gray's Oral Reading Test — How the tests are scored.

#### III. AN ESTIMATE OF READING TESTS.

Criteria for estimating values — These criteria applied.

#### IV. THE SERVICE OF READING TESTS.

Service to the superintendent — Reveals wrong emphasis in teaching — Service to the teacher — Service to the child — Remedying the situation revealed — Types of situations revealed — A normal situation — To raise the median score — Overemphasis on oral reading — Care from the beginning — Reading above the primary grades — Reading in the upper grades — Where variability is too wide — The difficult but normal case; suggestions for helping — Questions and topics for investigation — Bibliography.

## CONTENTS

### CHAPTER IV. SPELLING . . . . . 112

#### I. THE PROBLEM OF MEASUREMENT IN SPELLING.

Difficulties encountered — The foundation words of the English language — Making a spelling test.

#### II. SPELLING SCALES.

1. The Ayres Spelling Scale — How constructed — Pupils who are not tested — How many words to use — Methods of giving the test — Letters per minute — What the Ayres scale really is — Directions for giving a timed sentence test.
2. The Buckingham Spelling Scale.
3. Starch's Spelling Scale — Measuring the extent of the ability to spell — Starch's spelling lists.

#### III. STANDARDS.

Ayres's scale — The Starch tests.

#### IV. HOW TO LOCATE SPELLING DIFFICULTIES.

Locating bad spellers — Individuality in spelling difficulties — "Spelling Demons" — Types of misspellings — Teaching the pupil to correct his errors in spelling — Causes of some misspellings — Good teaching of spelling — Devices for improving spelling — Making associations automatic — Curtis's spelling practice tests — Questions and topics for investigation — Bibliography.

### CHAPTER V. HANDWRITING . . . . . 145

#### I. THE PROBLEM OF MEASUREMENT IN HANDWRITING.

Another method of measuring.

#### II. HANDWRITING SCALES.

Measuring speed — Selections for the speed tests — Measuring quality ; use of scales — The score card for detailed analysis — The scales classified as to use — Methods of using scales — Measurement for diagnosis — Use of the score card — The Freeman Scale — Using the Freeman Scale.

#### III. THE RELIABILITY OF MEASURES AND SCALES.

Rate and quality contrasted — Accuracy of the scores — Training in using the scales — Relative values of the different scales.

#### IV. STANDARD SCORES.

Freeman's proposed standards — What these standards represent — Other evidence as to standards — Standards required for work.

#### V. THE TEACHING SITUATION REVEALED.

Handwriting a complex ability — An individual rather than a class problem — Children differ widely in abilities

## CONTENTS

and needs — Plotting scores, and reading their meaning — Successive measurements to reveal progress — Meeting the situation revealed — Systems of penmanship — Movement in handwriting — Rhythm — Speed — Quality and speed — General laws of learning applied — Devices of remedial instruction — Increasing speed — Developing rhythm — Motivating practice — Reasons for using handwriting scales — Questions and topics for investigation — Bibliography.

### CHAPTER VI. LANGUAGE . . . . . 192

#### I. THE PROBLEM OF MEASUREMENT IN LANGUAGE.

One of measuring specific habits.

#### II. THE MEASUREMENT OF ABILITY IN ENGLISH COMPOSITION.

1. The Hillegas Scale.

2. The Harvard-Newton Composition Scale.

3. The Breed and Frostic Scale.

4. Willing's Scale.

5. The Nassau County Supplement.

Reliability of measurements — Use of the scales — Directions for using the Hillegas Scale — Hillegas Scale scores — Directions for using the Harvard-Newton Scale — The Harvard-Newton Scale scores — Directions for using the Willing Scale — The Willing Scale scores — The Willing Scale reproduced.

#### III. THE MEASUREMENT OF LANGUAGE ABILITY BY COMPLETION TESTS.

The Trabue Completion-Test Language Scales — Trabue test standards.

#### IV. THE MEASUREMENT OF ABILITY IN ENGLISH GRAMMAR.

Types of ability — Starch's Grammatical Scales — The Punctuation Scale — The Grammar Tests — Buckingham's Test.

#### V. MEASURING ACCURACY IN COPYING.

The Boston test — Kinds of errors made — Misspelled words — Undotted "i's" and uncrossed "t's."

#### VI. EDUCATIONAL SIGNIFICANCE OF THE USE OF THESE SCALES AND TESTS.

Finding specific language weaknesses — Remedying the situation revealed — Analyzing language ability — Questions and topics for investigation — Bibliography.

### CHAPTER VII. HIGH-SCHOOL SUBJECTS . . . . . 224

#### I. ALGEBRA.

The problem of measurement — The fundamental opera-

## CONTENTS

tions of Algebra — Standard Research Tests in Algebra — Other Algebra tests — Conclusions from the tests — Standards — Meeting the teaching situation revealed by Algebra tests.

### II. GEOMETRY.

### III. FOREIGN LANGUAGES.

Starch's language tests — The Hanus Latin Tests.  
Henmon's Latin Tests.

### IV. PHYSICS.

### V. OTHER TESTS WHICH MAY BE USED IN THE HIGH SCHOOL.

Questions and topics for investigation — Bibliography.

## CHAPTER VIII. STATISTICAL METHODS . . . . . 241

Good arrangement of scores — The median — Frequency of scores — Intervals of distribution — Approximate and true median — The average — The mode — Measures of variability: (1) Average deviation; (2) Percentiles; quartiles; probable error — Correlation — Coefficient of correlation — Graphical representation — Representing three qualities — Representing many qualities — Questions and topics for investigation.

## CHAPTER IX. THE MEANING OF SCORES . . . . . 258

Indefiniteness of school marks — School marks *vs.* scores — Translating school marks into scores — A satisfactory standard — An efficient standard — Effort to be expended on the tool subjects — School demands on the tool subjects — Basis for standards of accomplishment — Types of standards — Questions and topics for investigation.

## CHAPTER X. THE DERIVATION OF TESTS, AND EXAMINATIONS . . . . . 273

Analyzing pupils' class work — Bases for evaluating pupils' work — The cycle principle — The per-cent-of-pupils-solving basis — A normal distribution of ability — Points to be considered in evaluating exercises — Opinion of competent judges — The teacher-judgment basis — Reliability important.

Making examinations more effective — Care in framing questions — Questions and topics for investigation.

## CHAPTER XI. USE OF STANDARD TESTS IN THE SUPERVISION OF INSTRUCTION . . . . . 284

Assisting teachers — Four steps in supervising instruction with tests — Giving the tests — Tabulating the scores — Interpretation of the scores — Remedial treatment — Teachers need detailed and definite specifications — Courses of study represent working

## CONTENTS

specifications — Such subject-matter directions not quantitative — Such directions lead to formal and uniform instruction — The Tests aim to introduce quantitative work — Tests introduce scientific management — Handwriting an example of wasting time — The Cleveland Reading results a study in efficiency — Standards for instruction illustrated from Arithmetic — Results of using the Courtis tests in Boston — The supervisor and the standard tests — Questions and topics for investigation.

INDEX . . . . . 303



## LIST OF CHARTS AND FIGURES IN THE TEXT

1. Distribution of marks assigned to one Geometry paper by 116 teachers . . . . .	7
2. Distribution of scores with the Stone Reasoning Test in Butte . . . . .	43
3. A chart, showing the scores made by a sixth-grade pupil, in comparison with standard scores, using the Courtis Arithmetic Tests . . . . .	51
4. Median scores for all Cleveland schools, and five selected schools (Test D, division) . . . . .	52
5. The Ayres Spelling Scale . . . . .	113
6. Showing distribution of 91 pupils, according to the number of words spelled correctly . . . . .	117
7. A Section of the Thorndike Handwriting Scale . . . . .	149
8. Two sections of the Ayres Handwriting Scale . . . . .	150-151
9. Standard score card for measuring handwriting . . . . .	155
10. Individual record card, Freeman Scale . . . . .	160
11. Showing the distribution of scores in handwriting of a third-grade class . . . . .	177
12. Showing the distribution of scores in handwriting of a fourth-grade class . . . . .	177
13. Showing the distribution of scores in handwriting of a fifth-grade class . . . . .	177
14. Results of the Composition Test in Salt Lake City . . . . .	201
15. Showing correlation of grades at entrance to college and in the Freshman year . . . . .	253
16. Graphic representation of the standard scores for Starch's Spelling Tests . . . . .	254
17. Another form of graphic representation of the same standard scores as in Figure 16 . . . . .	254
18. Representing graphically the standard for handwriting . . . . .	255
19. Representing a pupil's score in several subjects . . . . .	256
20. Showing the meaning of school grades in terms of scores . . . . .	261