

The Century Education Series

Educational Measurements

AND THE

Class-Room Teacher

BY

A. R. GILLILAND

PROFESSOR OF PSYCHOLOGY, NORTHWESTERN UNIVERSITY

R. H. JORDAN

PROFESSOR OF EDUCATION, CORNELL UNIVERSITY

FRANK S. FREEMAN

ASSISTANT PROFESSOR OF EDUCATION, CORNELL UNIVERSITY



THE CENTURY CO.

New York & London

PREFACE

In presenting an addition to the list of books treating of the general subject of educational measurements, some definite justification must be offered. The authors of the present text feel that there is a twofold need which has been but partially met by the otherwise admirable books now before the public. The first is the need for a text which may be used as a handbook for guidance of the teacher in service; the second is the need for a class-room text adapted to the use of prospective class-room teachers.

As to the first, there seems to be a haziness in the minds of many class-room teachers in both elementary and secondary schools as to the use they may make of tests and scales in their own work. They are too often of the opinion that tests are primarily supervisory instruments, and so too difficult in operation and too abstruse in interpretation to be of any real aid to the individual teacher. A certain emphasis, then, should be put upon the fact that achievement tests are valuable instruments for the teacher to understand and use independently of, or in coöperation with, the supervisor. The teacher should understand that she may handle her own work more intelligently, with more successful results, in proportion as she makes the greater use of proper measures of her effort.

Again, there has been confusion in the terminology

of tests, which has given rise to many false notions among the rank and file of the teaching profession. The expression "psychological tests" has been used as a blanket phrase to cover every sort of measurement, and many teachers have not learned to discriminate between various types of psychological measures. Comparatively few teachers, to judge from reactions in representative summer school groups, can, for example, explain accurately the difference between intelligence tests and achievement scales. Thus an unfortunate situation holds which must be cleared up in order that the work of the clinician and supervisor may be done with assurance of understanding and resultant sympathy from the body of class-room instructors.

The second great need refers to two aspects of the work in classes in education in normal schools and colleges which deserve a rather different emphasis than has hitherto been given. On the one hand is the class of students preparing primarily for class-room teaching, who wish to study the technique and meaning of achievement tests at the same time that they are given practice in their use and evaluation. In the single semester frequently given to this work, they do not wish to include a consideration of the intelligence test or other psychological measures, and thus need a manual with the major emphasis upon the use of the achievement test. The study of the intelligence test will come in a separate course, as will the factors of supervision and administration as related to measurement.

On the other hand, we frequently meet a group within college departments of education, made up of undergraduates who are concerned more with general method

and subject matter of instruction than in the psychological phases of the teaching process. Unfortunately there are still a great many of this sort of students; and the teacher of general methods of instruction finds that they have no idea of the meaning or use of the tests, and that frequently they do not have a place in their undergraduate programs for such special courses. There is a place for a manual which may accompany such courses in general method, to give this knowledge of the various tests and their uses, and so to prevent the student from going into the work entirely ignorant of the great possibilities of such instruments. The authors have had the thought of working out a textbook which may be adapted to both types of undergraduate classes.

The present volume, then, is an attempt to meet this twofold need; in order to meet the varying objectives the idea has been kept steadily in mind of presenting the important concepts and methods of application of tests and scales in as simple and non-technical language as is consistent with sound practice in their employment. The order of topics and general arrangement have been planned from this point of view. It is hoped that the materials are so presented that the book will find its place as a manual for the teacher in service and a suitable part of the reading circle on the one hand, and as a class-room text for the normal school, college, and school of education on the other.

It is impossible to acknowledge all sources from which materials for this text have been drawn. Some materials have been taken directly; others have been included in modified form, while still other materials

have served only as suggestions for our purpose. In the subject of educational measurements the fund of materials is so great that original sources are often very difficult to discover. Yet the authors have attempted to acknowledge their indebtedness wherever possible.

In addition to acknowledgments to publishers of textbooks from which material has been quoted, special acknowledgments should be made to the many publishers and authors of the tests and scales from which quotations and illustrations have been freely taken. Chief among these are: The Public School Publishing Co., World Book Co., Bureau of Publications, Teachers College, Columbia University, Russell Sage Foundation and Mr. S. A. Courtis.

The authors are especially indebted to Miss Margaret Gessford of Washington, D. C., who painstakingly read the original manuscript and gave many valuable suggestions for its revision.

THE AUTHORS.

PREFACE TO THE REVISED EDITION

In the field of educational measurements, new tests frequently appear, and the older tests are being constantly subjected to scrutiny and experiment for the purpose of evaluation. This revision, therefore, has been made in the light of the contributions during the six years since the appearance of the first edition. Where justified the less significant of the older tests have been eliminated from the text; and new tests of demonstrated merit have been included. Furthermore, older tests have been re-evaluated in the light of recent experiment. Most of the chapters have been enlarged, particularly those dealing with secondary school subjects. The discussion of statistics and graphic method has been expanded, primarily from the point of view of interpretation. In the general discussion (Chaps. I, II, III, and IV), additions and modifications were made wherever the need was indicated by more recent investigation. Throughout the text, however, we have held to the point of view as indicated in the preface to the first edition.

THE AUTHORS.

December, 1930

CONTENTS

	PAGE
CHAPTER I. REASONS FOR EDUCATIONAL MEASUREMENTS . . .	3
Traditional methods of grading—the need for more adequate methods of grading— <i>selected references</i>	
CHAPTER II. WHAT CONSTITUTES A STANDARD MEASURE . . .	20
Factors involved in standard tests—differentiations from traditional measures—necessary specifications of the objec- tive measure—the criteria of a good test—different kinds of measures—mental tests— <i>selected references</i>	
CHAPTER III. THE PRACTICAL USE OF EDUCATIONAL MEASURES IN THE CLASS-ROOM	40
Introduction—uses of achievement tests—for purposes of comparison—for diagnosis of teaching—for evaluating methods of teaching—for diagnosis of classes—for diagnosis of individual pupils—for setting standards of achievement —for promotion—uses of intelligence tests—for class diag- nosis—for individual diagnosis—the combination of achieve- ment and intelligence test results— <i>summary—selected references</i>	
CHAPTER IV. REQUISITES FOR GIVING OBJECTIVE TESTS . . .	67
The technique of objective tests—administering the tests— scoring the papers—interpretation of the scores	
CHAPTER V. SPELLING	78
Early objective tests—some problems of spelling—Ayres Spelling Scale—Buckingham Extension of the Ayres Scale —Monroe's Timed Spelling Tests—Iowa Spelling Scale— Material of English Spelling, Jones—The Teachers Word Book, Thorndike—Morrison-McCall Spelling Scale— <i>sum- mary—materials needed—supplementary list of tests—se- lected references</i>	

	PAGE
CHAPTER VI. HANDWRITING	98
Problems in the measurement of handwriting—Freeman Analytical Handwriting Scale—Ayres Measuring Scale for Handwriting—Thorndike Scale for the Handwriting of Children in Grades Five to Eight—Gray Standard Score Card for Measuring Handwriting—Courtis Standard Practice Tests in Handwriting— <i>materials needed—supplementary list of scales—selected references</i>	
CHAPTER VII. READING	123
Importance and problems of reading—types of tests—Gray Standardized Oral Reading Paragraphs and the Oral Reading Check Tests—Pressey First Grade Word Reading Test and First Grade Reading Scale—Thorndike Visual Vocabulary Scale and the Test of Word Knowledge—Monroe Standardized Silent Reading Test—Thorndike-McCall Reading Scale—Burgess Scale for Measuring Ability in Silent Reading—Courtis Silent Reading Test—Gray Silent Reading Test—Haggerty Reading Examination— <i>materials needed—supplementary list of tests—selected references</i>	
CHAPTER VIII. ENGLISH LANGUAGE AND COMPOSITION . . .	158
The problem of measuring language ability—types of language tests—Starch Punctuation Scale—Starch's English Grammar Scales—Charters' Diagnostic Language Tests—New York English Survey Tests—Cross English Test—Trabue Completion Language Scales—tests of composition—Nassau County Supplement to the Hillegas Scale—other composition scales—Abbott and Trabue Exercises in Judging Poetry— <i>conclusions—materials needed—supplementary list of tests—selected references</i>	
CHAPTER IX. ARITHMETIC	182
Importance and problems of instruction in arithmetic—types of arithmetic tests—Courtis Standard Research Tests, Series B—Courtis Standard Practice Tests in Arithmetic—Compass Diagnostic Tests in Arithmetic—Cleveland Survey Arithmetic Tests—Woody-McCall Mixed Fundamentals—Monroe Diagnostic Tests in Arithmetic—Monroe's Standardized Reasoning Tests in Arithmetic—New Stone Reasoning Test in Arithmetic— <i>conclusions—materials needed—supplementary list of tests—selected references</i>	

CONTENTS

xiii

	PAGE
CHAPTER X. GEOGRAPHY	208
The problem of measuring in geography—Hahn-Lackey Geography Scale—Posey-Van Wagenen Geography Scales—Buckingham-Stevenson Place Geography Tests—Courtis Supervisory Tests in Geography—Branom's Diagnostic Tests in Geography— <i>conclusions—materials needed—supplementary lists of tests—selected references</i>	
CHAPTER XI. HISTORY AND CIVICS	222
The problem of measurement in history—Hahn History Scale—Barr Diagnostic Tests in American History—Harlan Test of Information in American History—Van Wagenen American History Scales—Pressey-Richards Tests in the Understanding of American History—Columbia Research Bureau American History Test—Brown-Woody Civics Test— <i>conclusions—materials needed—supplementary list of tests—selected references</i>	
CHAPTER XII. MUSIC AND DRAWING	246
Music as a special talent—Seashore Measure of Musical Talent—Courtis Standard Supervisory Tests in Music—Kwalwasser-Ruch Test of Musical Accomplishment—Thorn-dike Scale for the Merit of Drawing by Pupils Eight to Fifteen Years Old—Kline-Carey Measuring Scale for Free-hand drawing— <i>conclusions—materials needed—supplementary list of tests—selected references</i>	
CHAPTER XIII. SECONDARY SCHOOL MATHEMATICS	261
The problem of measurement in mathematics—Rogers Test of Mathematical Ability—Kelley Mathematical values Test Alpha—Hotz First Year Algebra Scales—Douglass Standard Diagnostic Tests for Elementary Algebra—Illinois Standardized Algebra Tests—Minnick Geometry Tests—Schorling-Sanford Achievement Test in Plane Geometry—Columbia Research Bureau Plane Geometry Test— <i>conclusions—materials needed—supplementary list of tests—selected references</i>	
CHAPTER XIV. SECONDARY SCHOOL SCIENCE	280
Problems in the measurement of science—Van Wagenen Reading Scales, General Science, Scales A and B—Downing Range of Information Test in Science—Grier Range of In-	

formation Test—Ruch-Popenoe General Science Test—
Dvorak General Science Tests—Powers General Chemistry
Test—Rich Chemistry Tests—Iowa Physics Test—Hughes
Physics Scale—Columbia Research Bureau Physics Test—
comparison of tests in physics—other tests in secondary
school science—*conclusions—materials needed—supplemen-
tary list of tests—selected references*

CHAPTER XV. FOREIGN LANGUAGES 299

Objectives in the study of foreign languages—American Coun-
cil Alpha Tests in French, German, Spanish, and Italian—
American Council Beta Tests in French, German, and
Spanish—Columbia Research Bureau Tests in French,
German, and Spanish—Ullman-Kirby Latin Comprehension
Test—White Latin Test—other Latin tests—*conclusions—
materials needed—supplementary list of tests—selected
references*

CHAPTER XVI. GENERAL ACHIEVEMENT TESTS 316

The place of the general achievement test—New Stanford
Achievement Test—Sones-Harry High School Achievement
Test—other general achievement tests—*conclusions—mate-
rials needed—supplementary list of tests—selected refer-
ences*

CHAPTER XVII. INTELLIGENCE TESTS 330

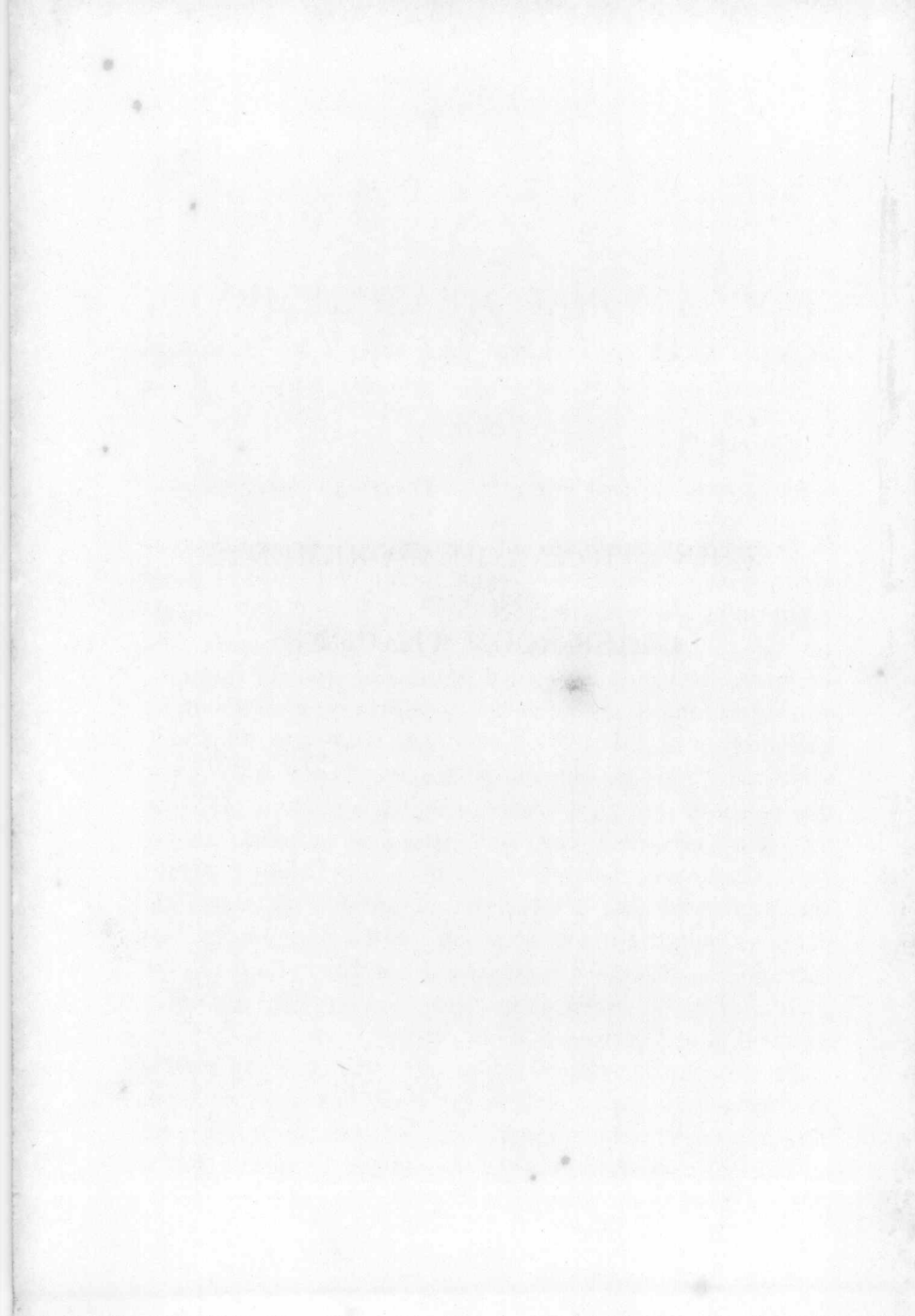
The nature of the intelligence test—types—some representa-
tive group tests—performance tests—aptitude tests—*mate-
rials needed—supplementary list of tests—selected refer-
ences*

CHAPTER XVIII. STATISTICAL AND GRAPHIC METHODS . . 346

Widespread use of statistics—statistical and graphic methods
—measures of central tendency—measures of deviation—
coefficient of variation—correlation—the column diagram
and the learning curve—frequency curves—*selected refer-
ences*

INDEX 391

EDUCATIONAL MEASUREMENTS
AND THE
CLASS-ROOM TEACHER



EDUCATIONAL MEASUREMENTS AND THE CLASS-ROOM TEACHER

CHAPTER I

REASONS FOR EDUCATIONAL MEASUREMENTS

Traditional methods of grading.—Teachers, more than most persons, have been privileged to pass judgment on others with respect to a very significant characteristic: namely, the extent of one's knowledge and learning. Not only is this an important quality in a person's total make-up, but it is one concerning which many individuals manifest a marked sensitiveness. This sensitiveness is found especially in parents who, of course, desire the knowledge and conviction that their children are of a high level. It is not altogether strange, therefore, that these parents have been unwilling to admit the correctness of a teacher's judgment of their children—a judgment expressed in terms of "marks" or "grades"—when the rating of the children does not coincide with their own conviction; and so the teacher is accused of unfairness and inability.

The teacher has arrived at her estimates by means of recitations and questioning, both oral and written. The answers to these questions she has evaluated; and on the basis of these values she has determined that a

certain pupil merits an "A", or a "B", etc. Such measures and estimates coming from the teacher are highly subjective; that is, they are the result chiefly of individual construction, opinion, and sometimes guess-work, wherein the personal equation has had more or less free play. Under conditions such as these it is inevitable that teachers' estimates have at times been in error, and that personal bias should have operated, with or without the awareness of the teacher.

It is only natural, therefore, that marked dissatisfaction with unreliable methods of interrogation and evaluation should have arisen not only among parents but among some teachers and other educators as well. Thus during approximately the last twenty years we have witnessed the development of measures of school learning known as objective subject-matter tests, the development being particularly pronounced during the last dozen years. These tests, as the term *objective* indicates, are intended to eliminate the influence of personal and individual bias and opinion in the measurement and scoring of school achievement. Yet, in spite of their widespread and increasing use, their nature and function are not clearly understood by all teachers, supervisors, and superintendents; nor by parents. It is appropriate, therefore, in opening our discussion, to review briefly the reasons for the construction of better educational measures than we have had in the past.

The need for more adequate methods of grading.—Three questions arise which must be answered in order to make these reasons clear: First, what need has the teacher for any concrete standards for measuring the work of the class-room? Second, why are the methods of

the past not adequate? Third, has any real progress been made toward a solution of this problem which may afford relief to the class-room teacher? These three questions are closely related and will be taken up in order.

THE TEACHER'S NEED.—The teacher needs some sort of concrete standard of measurement in at least three inevitable relations: (a) the need in relation to the public, particularly as represented by the parents of the pupils; (b) the need in relation to the superintendent of schools, the supervisor, the principal, or other officer in immediate charge of the teachers; (c) the need for standards in the relation with the pupils themselves, involving the teacher's own guidance in the class-room.

Need of standards in relation to the public.—The teacher must have some sort of means for acquainting the parent with the attainment of his child. Fathers and mothers expect this report as a matter of course, and if it is not forthcoming, insist upon it. Thus one reason for the failure of attempts to do away with school marks and marking systems has been the refusal of parents to agree that measures of progress are unnecessary. Doubtless the typical parent has been improperly educated in the application of such standards, especially to his own children, but this makes all the more important a kind of mark which will carry with it a clear idea of the reasons for the child's success or failure. Very few parents are content with a system which designates simply *Passed* or *Failed* as the teacher's verdict on the work of the month or semester. Parents generally wish to know something of the relative positions of their children; whether the children are working somewhere

near their mental capacity; whether they are diligent; whether they try to succeed; and something of the general character of their attempts.

The parent is not the only person interested in the work of the teacher. The mother's clubs, and other women's organizations, parent-teacher associations, the board of education, are all bodies which are interested in the work of the individual teacher as well as of the school system in general. The teacher is often asked to appear before one or another of these bodies, sometimes to explain the work of the class for the information of the public, sometimes on the defensive, to meet criticism, sometimes as a member of the body to extend the influence of the school. In such relations, easily explained and readily understood standards are most essential to give point and emphasis to her report. If she can show by definite comparisons that her work in such a subject as arithmetic, reading, history, or any other, surpasses the average as set by the national standard for that subject, she scores a distinct triumph and wins approval for her school and her city. Particularly is such a result a happy one if her teaching has been criticized or attacked; for when she has a standard with which to compare her work specifically, and to demonstrate her success without cavil, she can silence critics in the most satisfactory way.

Need of standards in relations with supervisory officers.—The work of the supervisor is to improve the results of the class-room teacher. This holds likewise for the relations of the superintendent, the principal, and the special supervisor to the teacher. Wise supervision is directed to the strengthening of the teacher,

and to capitalizing her strong points, not to the criticism of the instructor's weaknesses, or to an attempt to bring out her failures in bold relief. Her successes are to become habitual, her failures to become negligible. But in the past there has been much difficulty in the way. Teachers have claimed successes which the supervisor could not recognize for lack of an easily applicable measure of success. Superintendents have set up standards for judgment which the teacher could not understand; and, indeed, the supervisor has frequently not deigned to explain to the teacher the standard by which the judgment was made. One reason for this has been that the standard existed only in the mind of the superintendent and had not been defined in concrete terms in his own thinking. He might even go so far as to say that his standard was indefinable—that good teaching and bad teaching were to be *sensed* but that the difference could not be expressed in definite terms. Under such supervision, the teacher was largely helpless; she knew whether her work was approved or not, but she was not conscious of any standard by which this result was reached. So the need has been felt very definitely for standards by which the ideas of the supervisor might be passed on to the teacher in terms of objectives to be striven for by the teacher with a clear knowledge of the goal to be attempted; and by which there could be a clear understanding by both parties of the final decision as to success or failure to reach this goal. Such standards are essential if there is to be the sympathy between supervisors and teachers necessary to the best results. This sympathy has not obtained in the past in a large number of our communities; and there has grown up an