

McGRAW-HILL SERIES IN EDUCATION
HAROLD BENJAMIN, CONSULTING EDITOR

Educational Measurements
in the
Elementary School

McGraw-Hill
Series in Education

HAROLD BENJAMIN

CONSULTING EDITOR

The Stanford University Education Faculty—THE
CHALLENGE OF EDUCATION

Bobbitt—THE CURRICULUM OF MODERN EDUCA-
TION

Bowden and Melbo—SOCIAL PSYCHOLOGY OF
EDUCATION

Broom—EDUCATIONAL MEASUREMENTS IN THE
ELEMENTARY SCHOOL

Brubacher—MODERN PHILOSOPHIES OF EDUCA-
TION

Butts—THE COLLEGE CHARTS ITS COURSE

Carroll—GENIUS IN THE MAKING

Croxtan—SCIENCE IN THE ELEMENTARY SCHOOL

Grinnell—INTERPRETING THE PUBLIC SCHOOLS

Heck—THE EDUCATION OF EXCEPTIONAL CHILD-
REN

Horrall and Others—LET'S GO TO SCHOOL

Jones, Grizzell and Grinstead—PRINCIPLES OF UNIT
CONSTRUCTION

McKown—ACTIVITIES IN THE ELEMENTARY SCHOOL

McKown and Roberts—AUDIO-VISUAL AIDS TO
INSTRUCTION

Newlon—EDUCATION FOR DEMOCRACY IN OUR
TIME

Pringle—THE JUNIOR HIGH SCHOOL

Schorling—STUDENT TEACHING

Sears—CITY SCHOOL ADMINISTRATIVE CONTROLS

Sorenson—PSYCHOLOGY IN EDUCATION

Thorpe—PSYCHOLOGICAL FOUNDATIONS OF PER-
SONALITY

Updegraff and Others—PRACTICE IN PRESCHOOL
EDUCATION

Wert—EDUCATIONAL STATISTICS

Wilson, Stone and Dalrymple—TEACHING THE
NEW ARITHMETIC

Winslow—THE INTEGRATED SCHOOL ART PRO-
GRAM

Educational Measurements

in the

Elementary School

BY

M. E. BROOM, PH.D.
Assistant Superintendent of Schools
El Paso, Texas

FIRST EDITION
SECOND IMPRESSION

McGRAW-HILL BOOK COMPANY, Inc.

NEW YORK AND LONDON

1939

COPYRIGHT, 1939, BY THE
MCGRAW-HILL BOOK COMPANY, INC.

PRINTED IN THE UNITED STATES OF AMERICA

*All rights reserved. This book, or
parts thereof, may not be reproduced
in any form without permission of
the publishers.*

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

The developments of the past several decades have shown that some knowledge of educational measurements is an essential part of the equipment of the classroom teacher in the elementary school. Good testing has come to be closely correlated with good teaching. This volume attempts to give the historical facts and the philosophy necessary for an understanding of how educational measurements have developed, as well as to present information dealing with the construction and uses of achievement and mental tests. A consideration of statistics is included, since statistics is an essential tool in dealing with problems involving educational and mental tests, but this material has been kept at a minimum.

This work is intended as a guide both for classroom teachers and for students in teacher-training institutions. Technical discussions, such as are desired by technicians, have been reduced to a minimum; the criterion in choosing materials for inclusion in the discussions of different topics has been What does a teacher need to know? The discussions of measuring instruments in the different subject fields gives very few technical data; the author is convinced that those who wish such information will be served best by the information contained in articles in professional periodicals and in manuals accompanying standardized tests. Furthermore, an extensive study of standardized tests is necessary for good test construction. The treatment of mental tests is more technical than is that of standardized achievement tests, since teachers need specific information concerning these instruments, which they will use but which few will construct. The presentation of the materials has been guided in part by the nature of students' difficulties with this subject during the fourteen years the author has taught educational and mental measurements courses in teacher-training institutions, just as the content was selected to satisfy teachers' needs and interests, as evidenced during the past four years of the author's experience in a large public-school system.

The author's indebtedness to writers in the field is evident

throughout the volume. He wishes to acknowledge the courtesy of the many writers and publishers who gave permission to use certain of their published materials. Without this privilege, the work could not have been completed. Specific acknowledgments to authors and publishers appear at appropriate points throughout the volume. The author is indebted also to Beatrice H. Broom, who read the entire manuscript and gave numerous valuable suggestions in addition to checking many of the statistical calculations.

M. E. BROOM.

EL PASO, TEX.,
September, 1939.

CONTENTS

	PAGE
PREFACE	v
EDITOR'S INTRODUCTION	ix
CHAPTER I	
INTRODUCTION	1
CHAPTER II	
THE TERMINOLOGY OF EDUCATIONAL MEASUREMENT.	14
CHAPTER III	
THE TREATMENT OF STATISTICAL DATA	26
CHAPTER IV	
THE USE AND INTERPRETATION OF STATISTICAL MEASURES	55
CHAPTER V	
STANDARDIZED VERSUS TEACHER-MADE OBJECTIVE EXAMINATIONS	87
CHAPTER VI	
THE CONSTRUCTION AND USE OF TEACHER-MADE OBJECTIVE TESTS	105
CHAPTER VII	
STANDARDIZED TESTS IN THE LANGUAGE ARTS.	146
CHAPTER VIII	
STANDARDIZED TESTS IN ARITHMETIC	207
CHAPTER IX	
STANDARDIZED TESTS IN THE SOCIAL STUDIES.	225
CHAPTER X	
STANDARDIZED TESTS IN OTHER ELEMENTARY SCHOOL SUBJECTS.	236

	PAGE
CHAPTER XI	
INTELLIGENCE AND ITS MEASUREMENT.	253
CHAPTER XII	
INTELLIGENCE AND ITS MEASUREMENT: MENTAL TESTS . . .	275
CHAPTER XIII	
USES OF EDUCATIONAL MEASUREMENTS	308
APPENDIX A	
PUBLISHERS OF STANDARDIZED TESTS	319
APPENDIX B	
FORMS FOR THE COMPUTATION OF CERTAIN STATISTICAL CONSTANTS.	320
APPENDIX C	
SQUARES AND SQUARE ROOTS OF NUMBERS FROM 1 TO 1000 .	323
INDEX.	335

EDITOR'S INTRODUCTION

The modern progressive movement in education is sometimes attacked on the ground that it is opposed to the quantitative determination of a child's needs, capacities, and achievements. This attack is largely a process of belaboring a dead horse. Although it is true that certain reformers in the early days of the activity movement were suspicious—in many cases very properly suspicious—of some of the testing programs then in vogue, it is now clear that most modern educators are eager to employ every available instrument that will give precise information concerning the abilities of the learner and the progress of his learning. That is exactly the characteristic which makes them truly modern in their outlook on education.

The whole system of education as it is conceived in the most forward-looking schools requires the use of careful and exact measurement. Its methods of instruction are based on a recognition and utilization of individual differences which can be known only through measurement. Its guidance program is merely a slough of pious hope when it is not a morass of chicanery unless it rests squarely upon a foundation of measurement. Even the attainment of its highest aim—the development of those character and personality traits which are least susceptible to quantitative treatment at the present time—is aided rather than hindered by the use of properly constructed, applied, and interpreted measures of learning. The teacher today needs a more accurate, comprehensive, and critical knowledge of measurements than ever before in the history of his craft.

The present volume is designed to supply that knowledge. It is a compendium of the information needed by an elementary teacher in carrying out the measurement phase of his work. It describes clearly and succinctly the nature and uses of tests and examinations used in the elementary school. It gives a wealth of exercises for acquiring skill in the construction and application of teacher-made tests and in the use and interpretation of standard tests. It offers the reader an opportunity

to evaluate his own growing knowledge by a series of examinations on the book itself.

The author of this book is eminently qualified for the task of presenting his subject. For many years he has taught the principles and techniques of educational and mental measurement to prospective teachers in college and university. He has supplemented this experience with extended work in the day-by-day guidance of teachers in service. His scholarly command of the subject and his understanding of the everyday problems of workers in the field have enabled him to produce a volume of highest usefulness to the profession.

HAROLD BENJAMIN.

UNIVERSITY OF MARYLAND,
September, 1939.

EDUCATIONAL MEASUREMENTS IN THE ELEMENTARY SCHOOL

CHAPTER I

INTRODUCTION

Change is a characteristic of human civilization. Certainly change has occurred in the past, and change is now in progress. The changes that have occurred and are now occurring have brought and are bringing new demands on education, and, as a result, education itself must change to meet new demands. There is common agreement that schools must teach a greater variety of subject materials and that they must be taught even more effectively than subject matter has been taught heretofore. Children must be given a more comprehensive and thorough education than was available to their ancestors.

What Is Meant by Education?—The content presented to children in schools is the product of centuries of racial experience. It includes the products of human intelligence and of human folly, of human courage and human fears, of human successes and human failures. Education attempts to bring to students the best things within human knowledge: the best thinking of great philosophers; the techniques for and the products of abstruse and laborious mathematical calculations; the crystallization of emotion and thought, fixed in form, which are the productions of great artists—inspiring music and inspired words and beautiful masterpieces of sculpture and of painting—and the records of deeds of great men and of great nations. Education, then, may be thought of as a process of continuous and progressive and selective reconstruction of racial experience, functioning to bring to each maturing individual a knowledge and a command of those things which are involved in living a successful and an ideal existence.

In so far as society is concerned, education is the agency for transmitting the social heritage. Generation after generation has

learned those things which are included in the knowledge and in the skills that make up what is called our civilization. All that is included in human thinking and doing depends for its continuance upon this relearning by each generation. In so far as any individual is concerned, education is an attempt to adjust the maturing individual to the environment in which he lives by a well-balanced program of living.

The School as an Educative Agency.—It is well known that schools did not exist in primitive society; but education, adequate to the survival of the individual and of the social group, did exist. This was provided by the home and by the community. Man evolved the school as a substitute for these primitive educative agencies when life in the social group became so complex that parents were no longer able to devote sufficient time to their children, making it difficult for youths to secure an understanding and a mastery of the conditions of existence and skills in the circumstances of life and of action. Even today, the school is only one educative agent; youth continues to receive educative experiences from the life of the home and of the community, from the customs of the family and of the social group, from participation in industrial activities, from recreatory activities, from religious experience, and the like.

The school presents to the child primarily a mass of cultural material to be understood and mastered and some opportunity to acquire skills. It attempts to aid the youth in understanding the significant facts of past racial experience, in order that he may become a normal, useful adult member of the social group. In the academic sense, education may be thought of as a process of transmission from each generation to the succeeding generation of the knowledge and the skills that have been assembled through racial experience. It is not intended, obviously, that each individual shall master all of this knowledge and all of the skills; it is sufficient if these things are learned by some individuals, to preserve them for the use of the social group.

Measurement in Education.—Measurement in education involves the careful and scientific determination of each child's ability to learn, in terms of mental ability, vitality, interests, inclinations, ability to persist, and so on, and also the amount that he has learned at any given time. It is an important functional part of the school program, since its findings influence the

adjustment of the school organization, subject matter, and methods of instruction to the varying needs and abilities of pupils. The purpose of measurement in education is to provide an opportunity for each child to master the content of education for which he is ready and to check the efficiency of methods of instruction so that he may learn the subject matter without undue expenditure of time and energy and without error. It is obvious that modern education cannot function effectively without the use of a great deal of measurement.

Tests and Testing in Schools of the Past.—Some critics of current procedures using measurements in education seem to feel that tests have come only recently into school usage; that tests and examinations are a fad; and that, like other educational fads, measurements will disappear from the schools. This idea is fallacious. Tests and examinations have been the traditional forms of school measures since schools have existed. As early as 300 B. C., intellectual studies were a part of the curriculum of the College of the Ephebi in Athens, and there was some arrangement whereby examinations were held in certain of the intellectual studies, particularly in grammar, geometry, rhetoric, and music. Even earlier than this, students at the end of their apprenticeship gave an exhibition of their proficiency in arms, actually a performance test.¹

It is impossible in so short a discussion to trace the use of oral and written examinations in all schools through the centuries. It is certain that tests and examinations have been used continuously. Measurement had a definite place in the medieval universities; examinations were used at the University of Bologna during the eleventh and at the University of Paris during the twelfth century.² The use of examinations as a means of determining promotion from one school form (grade) to another is cited at the University of Paris during the sixteenth century, although it is apparent that standards were not always enforced rigorously. From 1395 to 1500, no candidate for a degree from the University of Paris was rejected.³

¹ J. W. H. WALDEN, *The Universities of Ancient Greece*, pp. 38, 135, New York: Charles Scribner's Sons, 1910.

² C. H. HASKINS, *The Rise of Universities*, pp. 17, 63, 122, New York: Henry Holt & Company, 1923.

³ H. C. BARNARD, *The French Tradition in Education*, p. 190 and footnote 1, p. 230, Cambridge, England: The University Press, 1922.

The Jesuits, during and after the sixteenth century, gave annual examinations which were carefully prepared and included both oral and written questions.¹ The Oratorian schools in the seventeenth century used the results of annual examinations, given in August of each year, in determining the promotions of pupils.² Oral examinations seem to have been used at Oxford University prior to the passage, in 1800, of the Public Examination Statute making tests mandatory.³ School examinations were important enough in the German secondary schools for official recognition and regulation, as evidenced in Prussia by the issuance of Regulations for the Leaving Examination in 1812.⁴ Perhaps enough has been cited to show that tests and examinations of some type have played a part in school procedures through the centuries and that often the results of such measurement have been significant in determining success or failure on the part of the student.

Measurement in Recent Education.—There has been no decrease in the use of tests and examinations in schools through the centuries; but during the past century, a change has taken place in forms of measurement. For oral questions and verbal answers have been substituted written examinations, including, in recent years, teacher-made, objective tests and standardized achievement tests.

The first recorded attempt to standardize measuring devices for classroom use was accomplished by the Rev. George Fisher about 1864 in the Greenwich Hospital School in England. This educator developed a scale book covering certain subjects taught in the school, providing answers of varying degrees of excellence to certain problem situations.⁵ His effort seems to have been lost to education for many years and to have had little or no subsequent influence.

¹ S. G. WILLIAMS, *The History of Modern Education*, p. 115, Syracuse, N. Y.: C. W. Bardeen, Publisher, 1896.

² BARNARD, *op. cit.* *The French Tradition in Education*, p. 171.

³ A. MANSBRIDGE, *The Older Universities of England*, p. 104, Boston: Houghton Mifflin Company, 1923.

⁴ F. PAULSEN, *German Education*, p. 212, New York: Charles Scribner's Sons, 1908.

⁵ E. B. CHADWICK, Statistics of Educational Results, *The Museum, A Quarterly Magazine of Educational Literature and Science*, 3: 479-484 (January, 1864).

Eleven years prior to the publication of the first standardized achievement test in the United States, Dr. J. M. Rice conducted an investigation of the spelling achievement of thousands of school children, using a list of 50 spelling words.¹ This was the first serious effort to secure uniform measurement in education in this country. Rice not only surveyed the spelling achievement of large numbers of pupils; he also analyzed his results and attempted to plan educational procedures based on his findings.

The significant impetus in the development of educational measuring materials did not come until 1904, however, when Thorndike published the first book dealing with educational measurement.² This volume provided educators with a knowledge of the procedures necessary for securing and evaluating better measures of academic achievement. The first standardized achievement test, in arithmetic reasoning, was devised by Stone and published in 1908.³ During the following years, other tests were standardized in arithmetic, penmanship, spelling, and other elementary-school subjects. It is perhaps worth noting that the early measuring instruments were prepared for use in subjects that are themselves objective, as arithmetic and spelling. Not until 1916 were standardized tests made available in high-school subjects.⁴ The great majority of the many hundreds of standardized tests that have been and are now available for use in almost all school grades and subjects have been prepared and standardized since 1920. Testing, a definite part of school procedure, recently has been extensively stressed and quite markedly modified. Many millions of copies of standardized achievement tests have been used in schools during the past two decades.

The Purpose of Measurement.—The purpose of measuring instruments is to yield objective, unbiased, and impersonal evidence concerning something that it is desirable to know. In

¹ J. M. RICE, *The Futility of the Spelling Grind*, *The Forum*, 23: 163-179, 409-419 (April, June, 1897).

² E. L. THORNDIKE, *An Introduction to the Theory of Mental and Social Measurements*, New York: Teachers College, Columbia University, 1904.

³ C. W. STONE, *Arithmetical Abilities and Some Factors Determining Them, Contribution to Education* 19. New York: Teachers College, Columbia University, 1908.

⁴ J. M. LEE, *A Guide to Measurement in Secondary Schools*, pp. 10, 11. New York: D. Appleton-Century Company, Inc., 1936.

so far as teachers and pupils are concerned, this something has to do with the learning of the pupils.

It is not the purpose of scientific measuring instruments in education to mechanize education or educators, leading to the development of a product as standardized as objects turned out by a machine. When tests are used by someone adequately trained for such work, the result is to ascertain with a minimum expenditure of time and money what each child can and cannot do. The effort thereafter is to adapt subject matter and instructional methods to the capacities and needs of individuals as well as of the class group. The net result, when the measuring is carefully done and the instruction is thorough, is to develop the individual far beyond the point to which he could be brought if his needs and capacities had not been carefully diagnosed and evaluated.

Measurement and Teaching.—One result of the extensive use of standardized tests during the past two decades has been a recognition that curriculums as well as pupil achievements differ. Educators have been forced to an awareness that different things are taught in schools in different geographical areas in the United States and that frequently when the same things are taught in the same grade levels, widely differing results obtain. As a result of these things, education has engaged in recent years in many curriculum studies and in many investigations in teaching methods in different school subjects and grades. Much of this represents attempts to adjust curriculums to the needs of children rather than to force children to fit existing curriculum patterns.

Education has seen some interesting efforts to meet the problem of individual differences in children, inspired indirectly, at least, by results of measurement. The present accent upon teaching through activity may be cited as an example of this. It is true that standardized tests do not prove helpful in many experimental situations; they seldom are satisfactory to measure achievement in any but the most conventional subject fields. For example, there are no tests devised to measure creative ability or research ability, nor are there any measuring instruments available to evaluate adequately the aesthetic or the abstract qualities of the products of creative and research skills, even at the level of the upper elementary grades.

Measurement and Guidance.—Each child needs educational guidance, even though he may be taught as one of many individuals in a group by relatively uniform methods. Educational guidance that is not based on information obtained by the proper use of measuring instruments verges on speculation and quackery. Using tests will relieve the pressure on the teacher of minutely studying each child, an impossible burden in a class group of average size given to a teacher usually for one semester only, assuming that the teacher attempts to guide the learning of her pupils in the subjects that they study as well as to give them educational guidance in general. Furthermore, the teacher can well use the time and energy saved through the use of tests in studying her pupils and in aiding them individually and collectively both academically and spiritually.

Some persons have urged that all this may be accomplished in a different fashion, using the judgments of experts instead of specific answer examinations as a basis for evaluating and treating pupil needs. There is no question that the judgment of an expert is valuable and that it usually contributes something that tests cannot yield, particularly when the teacher is considering problem pupils. Yet the judgment of an expert, in turn, cannot yield immediately all of the information that commonly is needed. Furthermore, very few teachers, and particularly young teachers, would care to declare themselves to be experts in the educational guidance of children. Even when the teacher is admitted to be an expert, her judgment plus the findings of adequate measuring instruments combine to give a safer basis for planning than either the judgment of the expert or the test data alone could give.

Measurement and the Development of Personality and Character Traits in Children.—The majority of educational tests yield information dealing with subject achievement only. Because of this, some persons choose to argue that testing places the emphasis of teaching on wrong aspects of the teacher's job. This means that constant watching of test findings concentrates the teacher's attention on seeing that the child learns that $2 \times 2 = 4$, and so on, and blunts her finer sensibilities and appreciations of the characters and personalities of her pupils to such an extent that she neglects spiritual values. In other words, the teacher becomes an automaton with no time or energy to devote to influencing the personality and character of her pupils. The issue

cannot be avoided that the school and the teacher must function in developing the character and personality traits of each immature individual enrolled in the school under the direction of a teacher. The school must aid, too, in developing the pupils' aesthetic appreciations and their concepts of value. It is true that most of this cannot be done through the use of scientifically constructed examinations, but tests help both by yielding necessary information concerning the child and by conserving the teacher's time in connection with certain necessary classroom activities.

Moreover, teachers teach quite as often by example as they do by precept when they deal with intangibles such as personality and character. Pointing out the lessons of history, which emphasize the need for honesty in dealing with ourselves and others, will fail to function if the teacher conducts her classroom work in history in ways that permit pupils to receive high marks without knowing the content of the course. In a classroom where the teacher is busy cultivating a "lovely atmosphere," careful analysis shows that the pupil is frequently busy impressing the teacher rather than in learning, *i.e.*, mastering subject matter. Usually, too, such pupils are well aware what they are doing. In such a situation, tests to compel subject-matter mastery would contribute directly to developing desirable rather than undesirable human traits. Unfortunately, few teachers know their pupils' needs in arithmetic, spelling, and the like, when they are concentrating on studying their social traits and other characteristics rather than their subject needs and are doing nothing to ascertain directly what their subject-matter needs may be. Directly and indirectly, then, the correct use of tests and the proper evaluation and use of findings that they yield will further a program for developing spiritual values through classroom teaching. There is little validity in the contention that scientific measurement inhibits any program for the development of personality and character traits in pupils, provided the measuring instruments have been properly used by persons adequately trained in their use.

Measurement and the Child.—Occasionally one encounters the objection that the child's mind is too complex and too subtle to be measured adequately and completely by measuring instruments. It is true that tests are not available to measure all

human reactions. For example, it is not yet possible to measure initiative and creative originality in a child or in an adult. Psychology has not yet provided measures of aesthetic appreciation of beautiful music and painting and poetry and sculpture.

Children, whether engaged in solving arithmetic problems, in designing an animated map dealing with manufacturing in the United States, or in playing a strenuous game of tag, are motivated by values that are more or less clearly recognized by them. These values, needs, wants, desires, and aspirations are wrought in the living fabric; yet psychological and educational measuring instruments ignore them. If a teacher seeks a measure of John Smith's ability to add $2 + 4$ to obtain 6, together with other number combinations, in every instance obtaining the true end result, she should be satisfied with this valid and accurate measure of this one thing. Tests tell *what*, not necessarily *why* or *how*.

There is in educational measurement a maxim known as the law of the single variable. This provides that a given test should, if possible, measure one thing only. If, then, a teacher wishes to know whether or not a given pupil can read and comprehend the meaning of the printed English language, she must apply a test of silent reading comprehension. If she wishes to know the probability of a given pupil's honesty, she must use one of the still relatively crude instruments devised to measure potential honesty.

Tests are not yet available for all purposes; with added research, better and more measuring instruments may be made available. Educators will then use them to advantage, measuring with each one the specialized study for which it was devised and validated. If, then, a general measure is desired, and time and finances permit, it will be requisite only to apply a multiplicity of tests and to strike a kind of average with the aid of statistics. Such a general average may have theoretical advantages, but it would be so complex that it would be of relatively little value in practical classroom situations. In the last analysis, what is desired is the child's perfect adjustment to his environment, meaning in so far as the classroom is concerned that he will learn effectively and adequately. This does not necessarily involve a complete scientific study of each individual child, yielding theoretical information as to the complexity or the simplicity of his personality and mentality. Tests provide, within their

limitations, the knowledge necessary for the teacher's use in aiding and guiding the child; this is all that should be demanded of them.

Measurement and Supervision.—Supervision is intended to provide the classroom teacher with assistance and direction so that she may obtain maximum results with her children with a minimum expenditure of time, effort, and money. The supervisor's program is essentially one of research and of service. Wherever supervision functions, measurement is of value, although teaching is so complex and involves so many factors for which objective measures are not available that supervision cannot be made entirely scientific.

When they are properly used, tests serve numerous purposes in supervision. In the first place, given early in a school term they serve to detect the most pressing subject-matter needs of class groups. They may also be used to analyze the subject-matter difficulties of individual pupils. Thus it is possible to determine whether a specific weakness affects an entire school population or is confined to one school, to certain grades in one or more schools, or to certain pupils only. This information gives an alert supervisor a starting point for guiding the teachers who work under her direction. It is her business to develop curriculums, remedial instruction programs, visual education materials, and teaching aids, such as drills, for the teacher's use in remedying pupil subject-matter weaknesses. If this remedial work is not devised properly, and if the supervisor does not check to see that it is properly applied in the classrooms under her direction, naturally no good results will follow testing procedures.

It is true that much of the information provided by tests can be obtained by alert supervisors and teachers through studying pupils intensively, but such study frequently consumes so much time that too little time is left for efficient instruction of the pupils who have been so studied. Objective measurement frequently discloses differing progress indices for class groups of apparent learning similarities. Frequently, teachers do not realize that class groups may be doing work either superior or inferior to their rating of it. Tests may yield information that makes it possible for the teacher, together with the supervisor, to provide opportunities for the pupils within their capacities and to see that full advantage is taken of such opportunities. Tests,

furnishing evidence of good and bad results, lessen the chance of weak teaching and increase the chances of good teaching throughout a school system.

Conclusion.—The purpose of the preceding discussion has been to show the function of education and of the school in the present civilization, as well as the function of measurement in modern schools. An attempt has been made to present briefly the highlights in the history of measurement in schools, past and present, and also to develop briefly an adequate philosophy of measurement as it functions in modern education.

There is no doubt that educators with a limited knowledge of measurement have misused measuring materials heretofore, and there is equally little doubt that the educator who casually accepts and uses carelessly standardized tests will encounter difficulties. Measuring is a scientific process, and it should be reserved to those adequately trained to use tests and examinations properly and to interpret correctly results that they yield.

The following chapters undertake to present the minimum essentials of the knowledge that a teacher should have to use tests and examinations properly in an elementary school. The discussion deals with terminology; with statistical devices and procedures and the interpretation of data yielded by statistics; with the construction and use of teacher-made, objective tests; and with the characteristics and uses of standardized mental and achievement tests.

Readings for Further Study

The History of Educational Measurements.

1. J. M. LEE, *A Guide to Measurement in Secondary Schools*. Pp. 1-15.
2. W. S. MONROE, *An Introduction to the Theory of Educational Measurements*. Chap. I. Boston: Houghton Mifflin Company, 1923.
3. C. W. ODELL, *Educational Measurements in the High School*. Chap. II. New York: D. Appleton-Century Company, Inc., 1931.
4. G. M. RUCH, *The Objective or New-type Examination*. Pp. 3-8. Chicago: Scott, Foresman & Company, 1929.
5. C. RUSSELL, *Standard Tests*. Chaps. II, III. Boston: Ginn & Company, 1930.
6. M. R. TRABUE, *Measuring Results in Education*. Chap. III. New York: American Book Company, 1924.
7. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Pp. 4-20. New York: Ray Long and Richard R. Smith, Inc., 1932.

8. C. WOODY and P. V. SANGREN, *Administration of the Testing Program*. Chap. I. Yonkers-on-Hudson, N. Y.: World Book Company, 1932.

The Philosophy of Educational Measurements.

1. M. E. BROOM, *Conflicting Philosophies of Educational Measurement*, *Kadelpian Review*, 10: 175-179 (January, 1931).
2. A. R. GILLILAND and R. H. JORDAN, *Educational Measurements and the Classroom Teacher*. Chap. I. New York: D. Appleton-Century Company, Inc., 1925.
3. C. A. GREGORY, *Fundamentals of Educational Measurement*. Chaps. II, V. New York: D. Appleton-Century Company, Inc., 1922.
4. W. A. MCCALL, *How to Measure in Education*. Chap. I. New York: The Macmillan Company, 1922.
5. C. RUSSELL, *Standard Tests*. Chap. I.
6. M. R. TRABUE, *Measuring Results in Education*. Chap. I.
7. M. J. VAN WAGENEN, *Educational Diagnosis and the Measurement of School Achievement*. Chap. XVIII. New York: The Macmillan Company, 1926.
8. W. L. UHL, *Some Neglected Aspects of Educational Measurement*, *Journal of Educational Research*, 27: 241-246 (December, 1936).

Questions and Exercises

1. Prepare a short paper (not over 600 words) on examination practices in schools prior to 1800.
2. Prepare an outline, arranged chronologically, of the major events in educational testing since 1860, with particular reference to standardized tests.
3. Discuss critically, but briefly, the contributions of one pioneer in the field of educational measurements.
4. Trace the trends, the significant sequences of events, etc., in educational measurements at different times during recent decades.
5. Identify a number of significant objections to the use of standardized tests and teacher-made, objective examinations in classroom procedures. Answer them.
6. List at least three quotations chosen during your reading on this topic that seem particularly meaningful to you. Give exact footnote references, using standard form. Write a brief analytical comment on each quotation.
7. What is meant by "neglected aspects of educational measurement"? Why are they neglected at the present time?
8. Write a paper (of 500 to 1,000 words) on the subject "Conflicting Philosophies of Educational Measurement."
9. What are the arguments of those who contend that it is not possible to measure the school product successfully? To what extent are these arguments valid?
10. To what extent are the results yielded by tests necessary for supervisory or administrative purposes?

11. List ways in which standardized tests might be used in improving teaching efficiency; ways in which such tests might detract from teaching efficiency.

12. Why, in your opinion, is it necessary for a student to know something of the fundamental thinking underlying educational measurements?

13. From what fields of human knowledge have testing procedures and practices been drawn?

14. Prepare a list of statements (not over 25), each not over 40 words, containing an adequate philosophy of educational measurements. If any of these are quotations, give the exact reference for each such quotation, using standard form.

CHAPTER II

THE TERMINOLOGY OF EDUCATIONAL MEASUREMENT

Terminology is the technical vocabulary of a given field of knowledge, so the terminology of educational measurements refers to the nomenclature, the *words* used in this subject to convey meaning. Significant and meaningful words are used in vain if one cannot comprehend them, so a discussion of terminology is necessary when one enters into a new field of knowledge. The purpose of this chapter is to present terms commonly used in educational measurements and to discuss and define them.

Scale.—A scale is a group of test items, or exercises, arranged in order of difficulty. Such a scale, including a relatively small number of items, arranged in a conventional scheme of difficulty intervals, may be used for comparing with the performance of a given pupil, say in drawing or handwriting. When a scale includes an extensive number of items, it may be the source from which items are taken for inclusion in tests. Scale value is the numerical value attaching to one item or exercise, which places it in its proper position in relation to other items or exercises in the scale with regard to the characteristics, such as difficulty, that are measured.

Test.—A test is an instrument including a number of items or exercises devised to measure pupil achievement. Achievement tests are those which measure pupils' mastery of subject matter taught in school. Standardized achievement tests include items that have been selected and checked empirically. Such tests are provided with exact procedures controlling the methods of administration and scoring and with norms¹ and data concerning the reliability and validity of the test. Teacher-made, objective tests are similar in form to standardized achievement tests, but ordinarily they are not so carefully constructed. (See Chapter VI for an extensive discussion of this topic.) Essay examinations, a third type of achievement test, include discussion exer-

¹ A norm is an achievement standard or representative value, usually an average score of the subjects employed in standardizing a test.

cises, factual questions, and the like, requiring that the student demonstrate not only mastery of subject matter but also a mastery of organization and of ability to express his answer to the given exercise or question.

A mental test is a standardized instrument devised and used for measuring mental ability. Its function is to establish ability-to-do, as opposed to that of an achievement test, which functions to demonstrate what a pupil has done.

An aptitude, or prognostic, test is one that has for its function the prediction of a pupil's status at some future time, usually in a given subject field. The best prognostic measures now available are given by mental tests.

A diagnostic test is one that yields detailed information concerning pupils' achievement in one or perhaps more than one relatively restricted fields. This type of test includes several of each type of problem at each level of difficulty, making possible the location of specific pupil weaknesses as a basis for remedial instruction. The theory is that when a pupil misses several opportunities to answer given types of exercises at different levels of difficulty, he does not know how to do that type of problem at any level of difficulty.

A survey test is usually composed of a variety of items or exercises in a single field of school instruction or, perhaps, in more than one such field. The function of the survey test is to yield a general or average measure of pupil achievement in the field or fields measured. Single divisions of such a test contain only exercises of a single type but of varying difficulty. This type differs from the diagnostic test by having one exercise only of each type and difficulty level. Since a pupil may miss a single exercise for reasons other than inability to do the item correctly, the survey test does not yield an indication of pupil weaknesses. The omnibus test is a specialized form of the survey test. This combines exercises or problems from a variety of fields, as in a test of general information. Survey tests are sometimes grouped into what is known as a battery of tests. A battery is a group of several tests, each in a different instructional field, given to pupils at one time or in a short period of time as part of a single testing program.

Many standardized achievement and mental tests are available in several forms. Form, usually used to mean duplicate or equiv-

alent form, means one of several similar or identical tests. Test items or exercises in duplicate forms are equivalent in difficulty and similar in number and in other characteristics. The materials of the test—its content—are likewise similar.

Types of Tests.—Tests are of various types. The rate, or speed, test includes materials of uniform difficulty. The score yielded by such a test expresses the number of exercises or problems that the pupil does within a given time. The weakness of such a test is that it does not measure the achievement of all pupils; it includes items either too difficult for the weak pupil or too easy for the advanced pupil, or both, and so it does not measure adequately either extreme in achievement.

The power test includes items arranged in order of difficulty, from very easy to quite severe. It is a scaled test, with the items arranged by graduated steps. Such a test gives a good measure of pupil achievement as to difficulty, but it does not usually yield a good diagnosis, unless it is presented with many items at each difficulty level.

The cycle test includes identical elements, of the same or equal difficulty, recurring at regular intervals throughout the test. Each group is one cycle. The pupils necessarily do all operations, even though they are unable to complete the entire test. This is the advantage of the cycle test, since segregation of items into separate divisions according to type might permit a pupil to avoid exercises of a type difficult for him. The spiral test uses the cycle form but is arranged in accordance with the difficulty of the items. It is a power test arranged in cycle form. The quality or accuracy test includes scales for comparison with pupil achievements, as in composition writing and silent reading achievement. In practice, usually several of these types are included in a single test.

Characteristics of Tests.—A good test must possess the following characteristics: validity, reliability, objectivity, graduated difficulty, variety of items or exercises, discrimination, freedom from coaching, and economy as to time and cost. These will be discussed individually in the following paragraphs.

Validity means the measure of the degree to which a test actually measures what it is planned to measure. A measuring instrument is said to be valid when it accomplishes the purpose

that it is intended to perform. Validity is the most important single characteristic of a test.

Validity may be established by studying the content of the test from the standpoint of its agreement with the particular curricular materials in connection with which it is to be used. When a test includes in its content materials commonly taught in the given subject field at the given grade level, and particularly when it includes those things commonly stressed in the best teaching practice in that subject field and school grade, it is possessed of some claim to validity. The opinion of experts constitutes evidence of validity also.

Validity may be determined statistically, in terms of a criterion, *i.e.*, principle, law, fact, or other standard for comparison. For example, scores yielded by a standardized test that already has been validated may serve as a criterion for determining the validity of a test in the same subject field and school grade that is being studied. If there is high agreement between the scores yielded by the two tests, the test of unknown validity is said to be valid because of the agreement of its scores with those yielded by the valid instrument. A test may be validated also by administering it to a group that has been expertly taught the curricular materials included in the test and to a group similar in other characteristics but thoroughly unfamiliar with the materials of the test. If the untrained group earn zero or approximately zero scores on the average, and the trained group earn high scores on the average, it is obvious that the test does measure what it purports to measure and that it is a valid test.

Reliability is synonymous with consistency or accuracy of measurement. Accuracy refers in a general way to freedom from error. When a test is administered twice within a short period of time, the resulting scores should be accurate and consistent for each individual responding to the test, if the instrument is reliable. In other words, scores yielded on the second administration must be equivalent to a high degree with those for the same individuals responding to the test on the first administration. Either the same form of the test or a duplicate and equivalent form thereof may be used for the retest. It is not essential that identical scores be yielded for given individuals on the test and retest but only that the two scores be equivalent, with a uniform

known relationship existing between them. If such a uniform known relationship between scores on two successive administrations of a given test does exist to a marked degree, the instrument is consistent in measuring whatever it does measure, and it is a reliable measuring instrument.

It should be noted that if a test is not reliable, it cannot be valid. If, however, a test possesses marked validity, it will have satisfactory to high reliability. It is possible to establish that a measuring instrument has high reliability without its validity being known, for the instrument may measure a characteristic or group of characteristics that cannot be isolated. In such an instance, the test has validity, but it has been impossible to establish a known relationship against given criteria.

A measuring instrument is said to be objective when different persons using it to measure the same thing secure similar results. A good test is objective as to giving and scoring; it possesses complete and precise directions concerning its administration and scoring and accurate scoring keys. Without these, it will yield subjective results, since different results will be secured by different persons, or by the same person at different times, when the instrument is used in attempts to measure the same thing. Objectivity of measurement is highly important, since without it the results of measurement of one group at different times or of different groups at the same time or on different occasions cannot be attempted successfully; without test objectivity, no assurance exists that scores will be comparable.

The items or exercises of a good test will be arranged in accordance with a graduated scale of difficulty, preferably with respect to one variable. A variable is any characteristic or trait that may exist in different amounts, and a good test attempts to test one thing only. Difficulty is the characteristic of test items that, when present in a large degree, causes a large number of incorrect responses and, present in a small degree, permits a large number of correct responses. The difficulty of the exercises in a good test will be such that every individual responding to it will be able to do at least a few exercises correctly. The difficulty of the items should then increase gradually until the last of the series of exercises challenges the most superior ability. The test should be so arranged that no individual responding to it makes a perfect score, since, if any one individual earns either a zero score or a

perfect score on the test, the ability of that individual has not been tested adequately.

A good test is characterized by possessing a variety of types of items or exercises. Mental ability and academic achievement are demonstrated not by the successful functioning of one specialized type of mental activity but by the power of the individual to meet adequately a variety of situations, making successfully the necessary mental adjustments involved in each. During the learning activities of pupils in schools, every series of problems is prefixed by complete directions and by trial exercises. With immature individuals, it is customary to give explicit directions and preliminary trial exercises preceding test items proper, to insure an understanding of the task to be done. This does not interfere with the reaction of the individuals who are concerned with the test in meeting the variety of items that it contains.

A test is said to possess satisfactory discrimination when the scores that it yields vary in accordance with known differences in pupil ability. A good test will differentiate diverse levels of pupil ability rather definitely from the inferior through the average to the very superior in terms of the scores earned on the test.

A good test will be relatively free from the effects of coaching due to practice and resulting familiarity with the test. Scores tend to increase on consecutive applications of the same or of duplicate and equivalent forms of an instrument because of such factors as familiarity with the test content, with the directions for doing the test, with the types of test exercises, and with methods for doing the required tasks. A good test will yield comparable results with given groups, holding the practice effects to a minimum.

Finally, a good test is economical as to both time and money. Time spent in school costs money, so a good test will be capable of administration in a minimum of time; furthermore, its cost will be low. The test content and the scoring keys also will be arranged to facilitate scoring, so that this may be accomplished in a minimum of time. Good tests do these things without loss of measuring efficiency.

Derived Scores.—A score is a description in numerical terms of a pupil's performance on a test. Derived scores are measures derived or computed from original or obtained scores.

The symbol C. A. refers to chronological or life age; M. A., to mental age; S. A., to subject age (subscripts such as r , s , and a refer to subjects, as reading, spelling, and arithmetic; S. A. r , for example, refers to subject age in reading); E. A., to educational age, the average of subject ages; and A. A., to achievement or accomplishment or attainment age, usually taken as synonymous with educational age. Mental age (M. A.) may be defined as the age of the average child having on a mental test the same score as that made by a given child who has responded to the given test. For example, if a fifteen-year-old child makes a score on a given mental test equal to the average score made by all twelve-year-old children, he is said to have a mental age of 12-0 (twelve years no months.) Subject age (S. A.) is obtained in the same manner as is M. A., excepting that it is yielded by a test in a given school subject field and not by a mental test. Educational age (E. A.), as indicated above, is the average of subject ages, or it may be obtained as is M. A., except that, in this instance, a battery of achievement tests is used and not a mental test.

A grade placement score (G. P. score, synonymous with B score) expresses achievement in terms of school grade and months completed in that grade. It is determined for a given child, as are the several derived age scores, in terms of the average achievement of children in a given school grade location. Like the age scores, not including C. A., G. P. scores are primary derived scores. Obtained scores and primary derived scores are quantitative statements.

Certain quotients are commonly used in educational literature. Occasionally R (ratio) is used instead of Q. Thus, A. R. means achievement ratio.

Probably the best known of the quotients is the intelligence quotient, I. Q. It is determined by the formula

$$I.Q. = \frac{\text{mental age (months)}}{\text{life age (months)}} \times 100$$

The value 100 is used to eliminate decimals. Other quotients are obtained similarly. S. Q., subject quotient, is the ratio of subject age to life age; E. Q., educational quotient, is the ratio of educational age to life age; and A. Q., accomplishment quotient, is the ratio of educational age to mental age. The I. Q., E. Q., and

A.Q. have been used by some educators as a basis for the classification of pupils, more or less inaccurately, as these measures are qualitative, not quantitative. Their function is mainly prognostic. For example, the intelligence quotient is chiefly useful for determining mental age at some time other than the present. It is for this reason that the constancy of the I.Q. is of importance.

The fallacy of the A. Q. is worth noting. Because of the common classroom practice of concentrating most of the teacher's aid and attention on the academically weaker pupils, such pupils tend to progress faster than their normal true rate. The more capable pupils, on the other hand, largely freed from the teacher's close supervision, frequently do no more than is necessary to maintain normal grade location and progress and drift toward mediocrity. As a result, the A. Q. measures the less capable pupils as above normal for their expectancy, whereas the superior pupils measure below their expectancy.

Pupil Performance.—By pupil performance is meant what a pupil does on a test; usually it is expressed in terms of a score. Performance must be distinguished from ability or capacity, which indicates what a pupil is able to do. There are three dimensions descriptive of a pupil's performance: (1) the amount, or, when produced under timed conditions, the rate of achievement; (2) the quality or accuracy of the performance; and (3) the level of difficulty at which the performance is given. Pupil performance is affected by two types of errors: (*a*) constant errors, which are peculiar to the test (lessening its validity), tending to be in the same direction and uniform in amount for all individuals responding to the test; and (*b*) variable, or chance, errors, differing in amount for the different individuals responding to the test.

Concluding Statement.—This chapter has attempted to give information concerning terms commonly encountered in a study of educational measurements, except statistical terms and symbols. These will be dealt with in the following two chapters. For terms not included in this chapter, the student is referred to the standard medical and psychological dictionaries.

Reference for Further Study

H. C. WARREN, *Dictionary of Psychology*. Boston: Houghton Mifflin Company, 1934.

Questions and Exercises

1. What is the relationship between E. Q. and I. Q.? Which is the more significant? Why?
2. What is the difference between a speed test and a power test?
3. What is the difference between a prognostic test and a diagnostic test?
4. How are the norms of standardized tests derived? Summarize the merits and the limitations of different types of norms, including the derived scores.
5. When a test has been standardized, is it justifiable to consider its derived scores permanent? Why? Is it justifiable to consider such scores (and norms also) equally valid for different geographical areas? Why?
6. What information must be given to describe completely a pupil's achievement performance?
7. What requirements should govern the construction of school achievement tests? What should be the characteristics of such tests?
8. What is a test? A scale?
9. How may a teacher determine the validity of an examination of her own construction? The reliability of such an examination?
10. Is the I. Q. a diagnostic measure of a child's capacity? Does this measure indicate which pupils in a given class need the teacher's attention and supervision?
11. Should a teacher expect to have an A. Q. for all pupils in all subjects? Why?
12. List the three significant differences between standard tests and teacher-made objective tests, taking into account the characteristics of a good test.
13. Indicate the several ways for determining the validity of a test, and illustrate each.
14. How would a teacher proceed to validate an objective test, if no standardized tests were available to serve as criteria?
15. How is a raw score yielded by a standardized test converted into a subject age? Into a grade placement score? How do these derived scores differ from each other?
16. Explain the meaning of duplicate forms and time allowances as applied to standardized tests.
17. Of what advantage is it to scale a test? What is meant by graduated difficulty scale?
18. How may a teacher determine the reliability of an objective test?

EXAMINATION ON TERMS USED IN EDUCATIONAL MEASUREMENT

I. Matching

Directions: Read each item in the right-hand column, and then find the item which matches it in the left-hand column. Record the number of this correct (left-hand column) item in the parentheses before the item in the right-hand column. You may need to record more than one number in each parentheses. If you do not know which one of the items in the left-hand

column is the correct one, skip it, and continue with the next one in the right hand column. When you have completed matching the items in Section I, begin with Section II; and when you have completed matching the items in Section II, begin with Section III. DO NOT GUESS! Work as rapidly and as accurately as you can.

- | | | |
|---------------------------|---------|--|
| 1. Validity | (_____) | The numerical expression of an individual's performance in terms of the unit used in scaling the test |
| 2. Item | | |
| 3. Derived score | (_____) | The smallest unit of test construction |
| 4. Group test | (_____) | The highest level of difficulty on which a pupil has responded to the exercises with a specified degree of accuracy |
| 5. Mental age | (_____) | A measure computed from the original measures yielded by a test |
| 6. Point score | (_____) | A test that can be given to a number of individuals at one time by one examiner |
| 7. Practice effect | | |
| 8. Variable | (_____) | A pupil's score on a mental test expressed in terms of age |
| 9. Power test | | |
| 10. Intelligence quotient | (_____) | A test in which the exercises are not arranged in order of increasing difficulty or vary in difficulty |
| 11. Raw score | (_____) | A score yielded directly by a test |
| 12. Difficulty score | (_____) | A scaled test |
| 13. Irregular test | (_____) | The increase of the scores on one trial over those yielded by a preceding trial of the same test when there has been no coaching between the two administrations of the test |
| 14. Rate score | | |
| 15. Accuracy | (_____) | The measure of a pupil's rate of work |
| | (_____) | The characteristic, or trait, that exists in different individuals in a group in differing amounts |

II

- | | | |
|---|---------|--|
| 1. B score | (_____) | A.R. |
| 2. Derived score | (_____) | A grade placement score |
| 3. Accuracy | (_____) | The age of the average child having the same score as that made by a given child under consideration |
| 4. M.A. | | |
| 5. Achievement age | (_____) | Transmuted score |
| 6. Rate of work | (_____) | Accomplishment age |
| 7. Degree of difficulty of the given task | (_____) | Dimensions describing pupils' performance |

- | | | |
|----------------------|-----|-----------------|
| 8. Practice test | () | Validity |
| 9. Constant error | () | S.Q. |
| 10. Difficulty score | () | Fore-exercise |
| 11. Aptitude test | () | Prognostic test |
| 12. Omnibus test | | |

III

- | | | |
|---------------------|-----|--|
| 1. Cycle test | () | The average of subject ages |
| 2. A. A. | () | Actual age in number of years a child has lived |
| 3. Achievement test | () | E.A. |
| 4. Spiral test | () | Of little significance as an index of a pupil's effectiveness as a learner |
| 5. Diagnostic test | () | A score synonymous with difficulty score |
| 6. Variable | () | A test that has for its function the prediction of a pupil's status in a given field at some future time |
| 7. E.Q. | () | A test in which identical elements of uniform difficulty recur at regular intervals throughout the test |
| 8. Prognostic test | | |
| 9. C.A. | () | A test having the items scaled according to difficulty, as well as the cycle arrangement of items |
| 10. Power score | | |
| 11. Rate test | () | A test that measures pupil's mastery of school subject matter |
| 12. Point score | | |
| 13. Criterion | () | A test making possible the location of specific weaknesses |

II. Completion

Directions: In each exercise below, write the word or words needed to complete the statement correctly

1. A _____ error is one that tends to be in the same direction for all members of a given group of pupils.

2. In a technical sense, a measure is said to be _____ when by comparison with some other measure of known reliability it is apparent that it is fairly reliable.

3. The E. A., the I. Q., and the S. R. are examples of _____ scores.

4. If two or more similar forms of a standardized test yield equal scores, they are said to be _____.

5. A _____ is a statement of the average achievement of pupils in a particular grade or, sometimes, of a given age.

6. A _____ test is one that has for its function the prediction of a pupil's status at some future time.

7. A measuring instrument is commonly said to be _____ if it fulfills the function that it is intended to perform.

8. The term _____ may be considered synonymous with crude score.

9. The term used in a general way to refer to freedom from error in a test is _____.

10. A pupil's score is a description of his _____.

11. The consistency with which a test measures whatever it does measure is known as _____.

12. The degree to which a test measures whatever it purports to measure is known as _____.

13. The term psychometric refers to the measurement of _____ in its broadest sense.

14. A _____ scale may be said to be any scale that makes use of scores computed in terms of points.

15. The characteristic of an exercise that when present in large degree causes a large percentage of incorrect responses is known as _____.

16. The term applied to any principle, law, fact, or other standard by which validity may be determined is _____.

17. A cycle test so arranged that there is an increase of difficulty in successive subtests or exercises is known as a _____ test.

18. A valid test must be a _____ test.

19. An _____ test is one that measures the pupil's mastery of subject matter taught in school.

20. The term variable error is synonymous with _____ error.

21. A _____ test requires the giving of problems of equal difficulty.

22. A pupil's _____ is his response to a test.

23. A test is _____ when different examiners, using it to measure the same thing, secure identical results.

24. Remedial instruction is predicated on results yielded by _____ tests.

25. An accuracy test is one of _____.

26. A test score that may be influenced by the personal opinion of the examiner is said to be _____.

27. A test that consistently estimates a pupil's ability from one time to another is said to have _____.

CHAPTER III

THE TREATMENT OF STATISTICAL DATA

There are few problems involving educational measurements in the consideration of which some acquaintance with the correct treatment of statistics is not necessary. Even though a classroom teacher is far from being keenly interested in statistical techniques and procedures, it is necessary for her to know how to understand, or, if necessary, to withstand, the statistics with which she will be confronted from time to time while dealing with problems of measurement.

Statistics means aggregates of related numerical facts, estimated according to reasonable standards of accuracy, assembled in a systematic manner, and placed in proper synthesis. It involves the collection, tabulation, analysis, and presentation of quantitative data. Statistical methods include the devices of analysis and synthesis by means of which numerical data are scientifically collected and used in the elucidation of quantitative, numerical facts too numerous or too complex for simple observation and interpretation. In other words, statistics provides the educator with a research tool, a keen technical aid, for extracting the significant truths concealed in masses of numerical statements.

Distribution				Distribution			
<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>
88	79	27	26	55	57	63	47
59	63	59	60	18	20	50	39
21	22	48	46	56	59	67	66
49	51	31	36	47	46	76	74
93	96	84	87	50	47	55	53
71	84	71	59	46	48	69	70
63	58	80	82	44	47	44	41
64	63	19	21	39	42	61	38
84	85	51	48	60	49	40	38
79	63	40	29	69	81	62	68
46	38	39	42	76	69	73	68
37	40	23	18	72	73	63	64
11	14	56	44				

For example, Distributions 1 and 2 each include 50 numerical facts, each a score or quantitative numerical statement concerning pupil achievement on an objective test. It is not possible for any person to observe these two distributions casually and to determine from such observation the facts as to the central tendency, or average, in the two distributions; their dispersion, or scatter, about the central tendency; the relationship between the paired series of measures (each score being paired with the corresponding score in the other distribution); or other significant facts included in the record. This is the reason why the teacher needs to know the computation procedures necessary in analyzing and summarizing the facts presented in the two distributions on page 26.

How Much Statistics?—Usually the teacher needs to know three types of measures that are necessary in describing analytically a mass of numerical data. The most important of these is a measure of central tendency, *i.e.*, the point within the range¹ and on the frequency scale about which the values in the distribution tend to group themselves. The second is a measure of the dispersion or variability of the values about this central tendency. The third indicates the degree of correlation or agreement existing between two given paired series of numerical facts. To this group must be added a fourth type, a critical ratio, McCall's experimental coefficient, if a classroom teacher is to have at her command the techniques necessary in dealing with instructional problems involving measurement. A fifth type of measure, dealing with reliability of computed constants, is less necessary for the classroom teacher. She is not so much interested in the probability that a given average or other statistical constant will recur subsequently at or near the same value as she is in the average for the given group of data under consideration. Measures of reliability affecting computed statistical constants have little practical interest for the teacher interested in mastering the use of tools to use in connection with her classroom problems. For this reason, any extensive discussion of such measures has been omitted from this discussion; the student is referred to any standard textbook in statistics for such information as may be desired concerning measures of reliability.²

¹ The range is the distance from the lowest to the highest value included in the collection of measures.

² For example, see M. E. BROOM, *Educational Statistics for Beginning*

The classroom teacher should also have a knowledge of two types of frequency graphs, the frequency polygon and the histogram; and of certain types of summary and comparison charts. The following pages are devoted to outlining the computation procedures to be followed in calculating these constants and to a discussion of the techniques to be utilized in preparing the different types of charts. Chapter IV deals primarily with the interpretation of these statistical measures and charts.

Tabulation and Classification.--Educational data usually are collected by single cases. They consist of scores, measurements, or other descriptive quantitative statements concerning a variable.¹ The simplest way to handle the measures included in an aggregate of numerical facts is to list them individually, *i.e.*, one by one. When a large number of values of a variable have been noted in the arbitrary order in which they happen to occur, however, the observations must be combined or grouped in a way that will facilitate an understanding of the significance of the frequency aggregate. This must be done before mathematical analysis and graphic representation of the facts can be undertaken.

Simple tabulation involves the preparation of a frequency array, or frequency distribution. This is an arrangement that groups the frequencies of the values of the given variable in ordered classes. Classification aids in making a distribution more readily comprehensible, and it makes mathematical computations relatively easy. Ordinarily, to insure mathematically correct computation findings, from 15 to 25 class divisions, or intervals, must be included in the frequency table. For economy of space, and to insure accuracy in computation through lessening the opportunity for error, from 7 to 10 intervals will be used in the sample problems used for purposes of illustration in this discussion.

In the process of formulating a frequency table, it is necessary first to determine the division, or interval, that will be used. To determine the width of intervals, one procedure involves the location of the lowest and the highest values of the variable

Students, pp. 109-119, New York: American Book Company, 1936.

H. E. GARRETT, *Statistics in Psychology and Education*. (2d ed.), pp. 198-250, New York: Longmans, Green & Company, 1937.

¹ A variable is a trait or quality existing in different definite amounts in different individuals.

included in the aggregate of measures. The difference between these values is then divided by the smallest and by the largest number of intervals that may be used.

In order to illustrate the procedure, it is necessary to tabulate a series of numerical values. Distribution 1 (page 26) will be used for this purpose. The lowest score in this collection of measures is 11; the highest, 93. These values designate the limits of the range. The difference, 82, divided by 10, the maximum number of classes to be considered here, yields 8 as a result, decimals being disregarded. Similarly, divided by 7, the minimum number of classes desired, the result is 13. Since ordinarily computation is simplified if the score distance taken for the class interval is an even divisor of 100, the interval in this instance will be 10, the only even divisor of 100 between 8 and 13. Using an interval of 10, the frequency tabulation for Distribution 1 appears as follows:

Class	Tabulation	<i>f</i>
90 -	1	1
80 -	1111	4
70 -	1111 11	7
60 -	1111 1111	10
50 -	1111 1111	9
40 -	1111 1111	9
30 -	1111	4
20 -	111	3
10 -	111	3
—————		<i>N</i> = 50

It should be noted that all intervals ordinarily are of equal value and also that the lower limit of each is divisible by the interval value itself. This procedure should be followed systematically, as it facilitates computation and reduces the possibility of error.

Accuracy is essential in the plotting of the original frequency tabulations, since a mistake in the preparation of a frequency distribution will invalidate the accuracy of all the subsequent statistical computations. The most common error in tabulating data consists in placing scores in incorrect class intervals. Great care must be exercised in making certain that all measures are accurately placed. The only satisfactory check on the accuracy with which the data have been tabulated is a retabulation of the entire collection of measures.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Chap. I. 88
2. H. E. GARRETT, *Statistics in Psychology and Education*. Pp. 4-9. 59
3. K. J. HOLZINGER and B. C. MITCHELL, *Exercise Manual in Statistics*. Chap. II. Boston: Ginn and Company, 1929. 21
4. T. L. KELLEY, *Statistical Method*. Chap. I. New York: The Macmillan Company, 1923. 49
5. J. E. WERT, *Educational Statistics*. Chap. II. New York: McGraw-Hill Book Company, Inc., 1938. 93
6. G. U. YULE, *An Introduction to the Theory of Statistics*. Pp. 75-84. Philadelphia: J. B. Lippincott Company, 1932. 71

Averages. *The Arithmetic Mean.*—The first natural step in reducing the values of a variable that are grouped in a frequency distribution to a more usable form is to determine an average. The mean, commonly termed the arithmetic mean, is the most widely known and the most frequently used of all the averages. It has the advantage of being quite easy to calculate, and it is so often used that it is seldom thought of as being one of the technical measures of statistics. Its simplicity makes it the most easily understood of all the measures of central tendency, although it has the disadvantage of leading untrained persons to think that the values of the variable of which it is the average are clustered closely about it.

The mean may be defined as the sum of the values of all the measures included in a frequency distribution divided by the total number of such measures ($M = \Sigma x/N$). It denotes the size that the numerical values would be if all were exactly alike in value while the total number of measures remained the same. The computation of the mean is simple, since all that is necessary is to sum the given values and divide by the number of such values. For example, using the data in Distribution 1, the computation of the mean is as illustrated, yielding a result of 54.96.

Certain modifications and "short-cut" methods are available for computing the arithmetic mean. They entail some loss of absolute accuracy, but it is so slight that for most purposes it is

	88
	59
	21
	49
	93
	71
	63
	64
	84
	79
	46
	37
	11
	27
	59
	48
	31
	84
	71
	80
	19
	51
	40
	39
	23
	56
	55
	18
	56
	47
	50
	46
	44
	39
	60
	69
	76
	72
	63
	50
	67
	76
	55
	69
	44
	61
	40
	62
	73
	63
	50)2748
	$M = 54.96$

more than compensated for by the greater ease of calculation. Such losses tend to decrease as the number of cases included in the mass of data increase.

Only one such short method for the finding of the mean will be demonstrated here. This requires the preparation of a frequency table; and for purposes of illustration (here and hereafter, in demonstrating the computation of other constants), the frequency table including the measures in Distribution 1, prepared in the previous division of this discussion, will be used, eliminating the column for tabulation.

Class	f	d	fd
90 -	1	4	4
80 -	4	3	12
70 -	7	2	14
60 -	10	1	10
50 -	9		$\Sigma + fd = 40$
40 -	9	-1	-9
30 -	4	-2	-8
20 -	3	-3	-9
10 -	3	-4	-12
$N =$	50		$\Sigma - fd = 38$
			$\Sigma fd = 2$
			$c = 0.04$
			$i = 10$
			$ci = 0.4$
			$AsM = 54.5$
			$M = 54.9$

The first step in calculating the mean by this method necessitates assuming one class to contain the mean. This assumed mean AsM is taken to be the mid-point of the assumed mean class. It is not necessary for the interval containing the assumed mean actually to be the one in which the mean is eventually found. Let it be taken in the class that contains the midmost case, for convenience in computing the median later. It is customary to set off the assumed mean class by drawing lines immediately above and below the class in which the assumed mean lies.

The steps in calculating the arithmetic mean by this short method may be stated in summary form as follows:

1. Select the assumed mean class, preferably the class containing the midmost value in the distribution; identify it in the frequency table by drawing lines above and below it.

2. Assign deviation d values to all classes, giving the assumed mean class a d -value of 0 and proceeding progressively from the assumed mean class in assigning other d -values. Deviation values for all intervals above the assumed mean class are designated as positive (+); below, as negative (-).

3. Determine the total magnitude of the "forces" operating about the assumed mean. To do this, multiply the frequency (f) in each class by its corresponding deviation value d , and enter the result under the fd column for that class. Add separately the total of the $+fd$ values and the total of the $-fd$ values; find the difference; and prefix to it the sign of the greater quantity. Enter this result as Σfd in the computation table, since it is the algebraic sum of the fd values.

4. Calculate the correction (c). To do this, divide Σfd by N . Since $N = \Sigma f$, the resultant is the interval distance (a value of d) between the assumed and the actual mean values.

5. Multiply c by the value of i (interval). By this process is found the score-value difference between the assumed and actual mean values.

6. Calculate the value of the assumed mean (add $\frac{1}{2}i$ to the lower limit of the AsM class, l). The AsM is the mid-point of the assumed mean class. In the sample problem, the assumed mean class was taken at 50. The value of 50, the lowest value in the 50 - class, actually extends from 49.5 to 50.499999+. The assumed mean, at the mid-point of the class, is found by adding one-half the interval value 5 to 49.5, the lowest value in the 50-class.

7. Add the resultant ci algebraically to the assumed mean to obtain the actual mean.

To check the accuracy of the result of such a computation, it is necessary only to take a different class interval as the assumed mean class, repeating the calculations. The mean will result as the same value, if the arithmetic is completed correctly.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 28-32.
2. R. W. BURGESS, *Introduction to the Mathematics of Statistics*. Pp. 80-85. Boston: Houghton Mifflin Company, 1925.
3. H. A. GREENE and A. N. JORGENSEN, *The Use and Interpretation of Educational Tests*. Pp. 127-133. New York: Longmans, Green & Company, 1929.
4. C. W. ODELL, *Statistical Method in Education*. Pp. 67-77. New York: D. Appleton-Century Company, Inc., 1935.
5. H. SORENSON, *Statistics for Students of Psychology and Education*. Pp. 40-56. New York: McGraw-Hill Book Company, Inc., 1936.

The Median.—The median is another measure of central tendency that is used to advantage often in summarizing a fre-

quency distribution. It is an average value which is determined easily, unless the distribution includes more than 100 cases, not grouped in a frequency table, and it is one that represents quite well all of the measures in a frequency distribution.

The median may be defined as the point on the linear scale of values on either side of which one-half the measures in the given frequency distribution fall.¹ The median is the computed location of the midmost value in the range of a given frequency distribution; it is not a score value, however, as is the mean. Since the median is located within the range of the distribution so that one-half of the measures fall above it and the other one-half below it, it represents the entire frequency distribution by virtue of its position.

The computation of the median of a frequency distribution may be expressed in formula form as follows:

$$Md = l + \frac{\frac{N}{2} - S_b}{f_m} \cdot i, \quad \text{or} \quad Md = u - \frac{\frac{N}{2} - S_a}{f_m} \cdot i$$

In these formulas, N is the total number of cases in the frequency series; S_b , the number of cases below the midmost class; and S_a ,

¹ From this definition it will be seen that the median is percentile 50, also called decile 5. Each decile includes one-tenth of the cases, included in the frequency distribution. The general formula for all percentiles is

$$Per_x = l + \frac{\frac{xN}{100} - S_b}{f_x} \cdot i, \quad \text{or} \quad u - \frac{\frac{(100 - x)N}{100} - S_a}{f_x} \cdot i$$

The term x designates the desired percentile. All that is necessary is to determine the class containing the last case included in the group of cases below the desired percentile value and substitute in the formula properly and compute for the given percentile value, following the steps as outlined for the median. For an explanation of symbols, see the discussion of the median; the symbols have the same meaning, but values substituted in the formula will be different for different desired percentile values. If percentile 10 is desired, using the 50 cases in Distribution 1, it is necessary to find and to use the class containing the fifth value from the bottom of the range. The $xN/100$ in the formula becomes $10 \cdot N/100$, which reduces to $N/10$ when the $l +$ formula is used. $\frac{(100 - x)N}{100}$ similarly becomes $9N/10$, when

the $u -$ formula is used. The formula adjustments for all other percentiles are made similarly.

the number of cases above the midmost class. The symbol f_m indicates the number of cases in the midmost class. The letter l stands for the lower limit of the midmost class; the letter u , for the upper limit. As before, i represents the interval.

The steps necessary in calculating the value of the median are as follows:

1. Assemble the ungrouped data into a frequency table.
2. Divide the total number of cases by two, and then count from either end of the range by totaling the frequencies in the several classes until the midmost class has been identified.
3. Total the number of cases below and above the midmost class (use S_b in calculations involving the $l+$ formula; S_a , with the $u-$ formula).
4. Substitute in the formula the values of l (or u), N , S_l (or S_a), f_m , and i .
5. Divide N by 2, and subtract from the result the value of S_b (or S_a). Multiply this result by i , and divide this result by f_m . Add this result to the value of l (or subtract it from u , depending on whether the $l+$ or the $u-$ formula was used). The resulting product is the value of the median. Either formula will yield the same final result for a given frequency distribution, provided there are no errors in arithmetic.

The formula solution for the value of the median of Distribution 1, the illustrative problem, is as follows:

$$\begin{aligned}
 Md &= 49.5 + \left(\frac{\frac{50}{2} - 19}{9} \right) \cdot 10 & Md &= 59.5 - \left(\frac{\frac{50}{2} - 22}{9} \right) \cdot 10 \\
 &= 49.5 + \left(\frac{25 - 19}{9} \right) \cdot 10 & &= 59.5 - \left(\frac{25 - 22}{9} \right) \cdot 10 \\
 &= 49.5 + \frac{6 \cdot 10}{9} & &= 59.5 - \frac{3 \cdot 10}{9} \\
 &= 49.5 + 6.7 = 56.2 & &= 59.5 - 3.3 = 56.2
 \end{aligned}$$

The foregoing discussion has assumed that a continuous series has been considered. One special case should be noted. If the total number of cases in the given distribution is even, and one-half of this number of cases lies in and above a given class, with the remaining one-half of the measures in the classes below the given class, the calculation of the median is not necessary, since the median will be the lower limit of the class in and above which one-half of the total number of cases fall.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 33-36.
2. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 85-89. Boston: Ginn and Company, 1928.

3. C. W. ODELL, *Statistical Method in Education*. Pp. 77-87.
4. H. SORENSON, *Statistics for Students of Psychology and Education*. Pp. 70-85.
5. L. L. THURSTONE, *The Fundamentals of Statistics*. Chap. XI. New York: The Macmillan Company, 1925.

The two averages, the computation of which has been discussed in the preceding pages, will suffice to describe the central tendency of frequency distributions yielded by measuring instruments commonly used in evaluating classroom situations. It is apropos therefore to consider measures of scatter, or dispersion, of the frequency aggregate about the average.

Measures of Dispersion. *The Standard Deviation.*—The standard deviation is a measure that has been widely used in reporting how much a group of items tends to scatter from the central tendency of a collection of measures. It may be defined as the arithmetic mean of the square root of the squared deviations of values of a variable from their arithmetic mean. It is associated with the mean; a characteristic of the mean is that the sum of the squares of the deviations from it is a minimum. One standard deviation from the mean, taken in both directions, positively and negatively, includes approximately the middle two-thirds of the cases in the given frequency distribution (actually the middle 68.27 per cent approximately in a normal distribution of frequencies). A standard deviation of relatively small numerical value indicates that the frequency distribution from which it was derived is quite homogeneous, whereas a large standard deviation in the same manner indicates that the frequency distribution is heterogeneous.

The task of calculating the standard deviation exactly, using the formula $\sigma = \sqrt{\Sigma d^2/N}$, is laborious, and it provides great opportunity for arithmetical error. It requires (1) finding the exact value of the mean, (2) finding the exact deviation of each individual measure from the actual mean, (3) squaring each such deviation value and then finding the sum of the squared deviation values, (4) dividing the sum of the squared deviations by the number of cases, and (5) taking the square root of the result obtained by dividing the sum of the squared deviations by the number of cases to ascertain the value of the standard deviation.

Just as it is possible to compute a mean by a short method which yields almost the same value as the mean computed exactly,

so it is possible to compute the standard deviation by a short method and to obtain a similarly accurate result. The computation of the standard deviation for the items in a frequency distribution grouped into a frequency table requires the use of the following formula:

$$\sigma(SD) = \sqrt{\frac{\Sigma fd^2}{N} - c^2 \cdot i}$$

Distribution 1 again provides the data used in illustrating the calculation of the standard deviation by short method. It should

Class	<i>f</i>	<i>d</i>	<i>fd</i>	<i>fd</i> ²
90-	1	4	4	16
80-	4	3	12	36
70-	7	2	14	28
60-	10	1	10	10
50-	9		+40	
40-	9	-1	-9	9
30-	4	-2	-8	16
20-	3	-3	-9	27
10-	3	-4	-12	48

$$N = 50 \qquad \qquad \qquad -38 \qquad 190$$

$$\Sigma fd = 2$$

$$c = 0.04$$

$$\sigma = \sqrt{\frac{190}{50} - 0.04^2 \times 10}$$

$$\sigma = \sqrt{3.8 - 0.0016 \times 10}$$

$$\sigma = \sqrt{3.7984 \times 10}$$

$$\sigma = 1.95 \times 10 = 19.5$$

be noted that this involves most of the work required to obtain the value of the mean by short method. The steps necessary in solving for the standard deviation by short method, assuming that the mean previously has been found by short method, are as follows:

1. Determine the value of fd^2 for each interval in the frequency table (multiply d by fd , and check by multiplying f by d^2 .)
2. Total the values of the fd^2 resultants for the several intervals, to obtain Σfd^2 .
3. Divide Σfd^2 by N .
4. Square the value of c , the correction used in calculating the mean.

5. From the resultant of $\Sigma fd^2/N$, subtract the value of c^2 .
6. Extract the square root of the resultant of subtracting c^2 from $\Sigma fd^2/N$.
7. Multiply this square root value $\sqrt{\Sigma fd^2 - c^2}$ by the value of the interval, thus securing the numerical value of the standard deviation.

This calculation is carried through in terms of class interval values until the final step, at which point the final result, the value of the standard deviation, is reduced to original unit values.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 61-65.
2. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 108-110.
3. C. W. ODELL, *Statistical Method in Education*. Pp. 123-131.
4. H. SORENSON, *Statistics for Students of Psychology and Education*. Pp. 138-146.
5. G. U. YULE, *An Introduction to the Theory of Statistics*. Pp. 134-144.

The Quartile Deviation.—The quartile deviation is used to measure dispersion about the median in the same manner that the standard deviation measures dispersion about the mean. All that is necessary to ascertain the value of Q , the quartile deviation, is to compute percentiles 25 and 75, known as quartile 1 (Q_1) and quartile 3 (Q_3). The formula is

$$Q = \frac{Q_3 - Q_1}{2}$$

One quartile deviation distance, taken on either side of the median, embraces the middle half of the cases in a normal distribution of frequencies. Since not all distributions are normal, particularly when numbers of cases are small, as is the case with class-size groups, and since Q is not based on the actual values of all the measures (since it is based instead on percentiles), it is not so good a measure of dispersion as it is σ . Ordinarily, Q should not be used as a measure of dispersion when it is possible to compute σ .

Readings for Further Study

1. D. H. COOKE, *Minimum Essentials of Statistics*. Pp. 32-35. New York: The Macmillan Company, 1936.
2. H. E. GARRETT, *Statistics in Psychology and Education*. Pp. 34-38.
3. C. W. ODELL, *Statistical Method in Education*. Pp. 105-109, 117-118.
4. J. E. WERT, *Educational Statistics*. Chap. III.

The measures of central tendency and of dispersion which have been described in the preceding pages permit general descriptive comparisons of two or more distributions of a given variable, but they yield no information as to the agreement that may exist between paired measures (two measures for each individual included in the group of cases) in the aggregates of items. Since this information is frequently necessary for a consideration of problems involving measurement, it is necessary for the student to know something of correlation.

A Measure of Relationship. *The Coefficient of Correlation.*—Essentially, correlation is the study of paired numerical facts. The coefficient of correlation, which is designated by r , is the most frequently computed measure of relationship. Using the term in its general sense, it may be defined as the numerical index that compares and summarizes the degree to which the corresponding items in two paired frequency series depart from their respective tendencies. The coefficient is an index of straight-line or rectilinear relationship; and when the relationship between two variables is rectilinear, the coefficient measures the proportion of the variation of one variable which is associated with a second variable. It is, therefore, a measure of the relative significance of the conjoint variation in the two variables. The limits to the value of the coefficient of correlation lie between positive and negative unity, with the numerical values of r ranging from -1.0 through 0 to $+1.0$, *i.e.*, from perfect negative correlation (exact inverse relationship) through neutral agreement to perfect positive agreement.

A number of formulas have been derived for calculating this measure. Since, ordinarily, the class group includes from 25 to 40 individuals, it is possible to compute coefficients of correlation which ordinarily will be encountered, by the Pearson product-moment formula, utilizing the scattergraph. This is the procedure probably most often used in calculating the coefficient of correlation. The formula is

$$r = \frac{\frac{\sum xy}{N}}{\sigma_x \cdot \sigma_y}$$

When the computation is done by short method, involving the use of the scattergraph, or scatter diagram, the formula becomes

$$r = \frac{\frac{\sum xy}{N} - c_x c_y}{\sigma_x \cdot \sigma_y}$$

In the formula, the terms x and y are used to designate the deviations of the measures from the means of the two series.

The first step in computing the coefficient of correlation is the plotting of the scatter diagram. Heretofore, the preparation of the frequency distribution for illustrative purposes has involved one collection of frequencies, Distribution 1. Now it is necessary to plot a two-way frequency table, involving Distributions 1 and 2. What is done may be thought of as plotting Distribution 1 as a frequency table, as was done previously, then placing on the horizontal axis a second classification for Distribution 2 and moving the cases in each class (or row in the table) until each case is in its proper class in the frequency table for Distribution 2 as well as its proper class in the frequency table for Distribution 1. Each entry in the scatter diagram, then, indicates the placement of one pair of scores, one of the pair being in Distribution 1, and the other in Distribution 2.

The scatter diagram showing the results of plotting Distributions 1 and 2 in this fashion follows.

Distribution 2

Distribution 1	Class	10 -	20 -	30 -	40 -	50 -	60 -	70 -	80 -	90 -
	90 -									
80 -								1	111	
70 -						1	111	11	1	
60 -				1	11	1	1111	1	1	
50 -				1	111	111	11			
40 -			1	11	111	1				
30 -				1	111					
20 -		1	11							
10 -		1	11							

Distribution 1

Distribution 1	Class	10-	20-	30-	40-	50-	60-	70-	80-	90-	$T(f)$	d_y	fd_y	fd^2_y	Σx	Σxy
	90-									1	1	4	4	16	4	16
	80-							1	3		4	3	12	36	11	33
	70-					1	3	2	1		7	2	14	28	10	20
	60-			1	2	1	4	1	1		10	1	10	10	5	5
	50-			1	3	3	2				9					
	40-		1	2	5	1					9	-1	-9	9	-12	12
	30-			1	3						4	-2	-8	16	-5	10
	20-	1	2								3	-3	-9	27	-10	30
	10-	1	2								3	-4	-12	48	-10	40
$T(f)$	2	5	5	13	6	9	4	5	1	$N = 50$		Σfd_y 2	Σfd^2_y 190		166	
d_x	-4	-3	-2	-1		1	2	3	4							
fd_x	-8	-15	-10	-13		9	8	15	4		Σfd_x -10					
fd^2_x	32	45	20	13		9	16	45	16		Σfd^2_x 196					
Σy	-7	-15	-3	-9		10	8	12	4							
Σxy	28	45	6	9		10	16	36	16	166						

Σxy must check

$$c_y = \frac{\Sigma fd_y}{N} = \frac{2}{50} = 0.04 \quad c_x = \frac{\Sigma fd_x}{N} = \frac{-10}{50} = -0.2 \quad c_x c_y = 0.04 \times -0.2 = -0.008$$

$$c^2_y = 0.04^2 = 0.0016 \quad c^2_x = -0.2^2 = 0.04$$

$$\Sigma^2_y = \frac{\Sigma fd^2_y}{N} = \frac{190}{50} = 3.8 \quad \sigma^2_y = \Sigma^2_y - c^2_y = 3.8 - 0.0016 = 3.7984 \quad \sigma_y = 1.95$$

$$\Sigma^2_x = \frac{\Sigma fd^2_x}{N} = \frac{196}{50} = 3.92 \quad \sigma^2_x = \Sigma^2_x - c^2_x = 3.92 - 0.04 = 3.88 \quad \sigma_x = 1.97$$

$$\sigma_x \sigma_y = 1.95 \times 1.97 = 3.8415 \quad \frac{\Sigma xy}{N} = \frac{166}{50} = 3.32$$

$$r = \frac{\frac{\Sigma xy}{N} - c_x c_y}{\sigma_x \sigma_y} = \frac{3.32 - (-0.008)}{3.8415} = \frac{3.328}{3.8415} = +0.866$$

The steps in computing the coefficient of correlation, using the Pearson product-moment formula, based on the scatter diagram, are as follows:

1. Scatter the cases in their respective classes in the two-way classification, each entry representing the location of two measures (or one pair of measures, one in each frequency distribution).

2. Total each array (row and column), and sum the totals of each array. This total must check for both distributions.

3. Establish the midmost classes in both classifications, and plot the deviations of all classes from these two midmost classes. (This duplicates the procedure for finding the mean by short method; this duplication of the method for finding the mean and the standard deviation continues in steps 4 and 5.)

4. Compute fd_y and fd_x for each class; then Σfd_y and Σfd_x .

5. Compute fd_y^2 and fd_x^2 for each class; then Σfd_y^2 and Σfd_x^2 .

6. Compute Σx (actually Σfd_x) for each array separately (each row is itself an individual frequency distribution) by multiplying the frequency in each class in the given row by its corresponding deviation (e.g., in the dx row in the illustrative problem). Total the products ($d \times fd$) for all the classes, adding algebraically. The resultant is entered in the Σx column for that array.

7. Multiply Σx in each row by the d_y value of the same class (row) to obtain Σxy for that array. Total the Σxy values algebraically.

8. Compute Σy similarly (follow Step 6, using column arrays instead of row arrays), and determine Σy for each column. Multiply Σy by d_x in the same column to obtain Σxy . Total the Σxy values. The resultants of Steps 7 and 8 should be identical if no arithmetical error has been made.

9. Substitute in the formula, as shown in the sample problem on p. 41, and complete the computations.

Those who master and understand this one method for finding the coefficient of correlation can meet the requirements of the majority of classroom problems involving measurements. Research workers will recognize that this degree of mastery of correlation may lead into serious error, so it is recommended that persons investigating complicated problems of relationship study extensively the literature concerning correlation.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 75-86.
2. D. H. COOKE, *Minimum Essentials of Statistics*. Pp. 114-153.
3. H. E. GARRETT, *Statistics in Psychology and Education*. Pp. 251-288; 359-407.
4. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 141-154.

5. T. L. KELLEY, *Statistical Method*. Pp. 151-185, 221-223.
6. C. W. ODELL, *Statistical Method in Education*. Pp. 143-236, 250-259.
7. H. SORENSON, *Statistics for Students of Psychology and Education*. Pp. 138-146.
8. G. U. YULE, *An Introduction to the Theory of Statistics*. Pp. 157-209.

It should be noted that when the Pearson product-moment formula is used in calculating the coefficient of correlation, it is not necessary to compute the means and standard deviations separately. The information is available from the scatter-diagram computations: All that is necessary is to multiply each correction c by i and add the resultant to the mid-point of the midmost class, the AsM . To obtain the standard deviation, merely multiply the obtained σ value from the scatter-diagram computation by the appropriate interval value.

Frequently it is not enough to summarize two distributions by determining their ranges, their central tendencies, and their dispersion. If the two are dissimilar, what is the certainty that this dissimilarity may again be anticipated with other similar distributions of the same variable? The problem is: Is the obtained difference one such as may always be anticipated, or is it one such as may occur due to chance or to random sampling? Critical ratios are used to provide information concerning this problem.

A Critical Ratio. *The Experimental Coefficient.*—This measure was devised by McCall for use in evaluating the reliability or the unreliability of a given difference between means for the same variable. Since the formula is

$$EC = \frac{M_x - M_y}{2.78 \sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}}}$$

the experimental coefficient is a critical ratio based upon measures describing two series of numerical facts. In this formula, x is used to designate the distribution having the mean of larger value; y , that having the mean of smaller value. The values of the mean and of the standard deviation and the numbers of cases in both distributions are necessary in obtaining the experimental coefficient.

The steps necessary in obtaining the experimental coefficient are as follows:

1. Subtract the value of the smaller from that of the larger mean.
2. Square σ_x and σ_y .
3. Divide σ_x^2 by N_x , and σ_y^2 by N_y . Add the two resultants.
4. Take the square root of the resultant of $\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}$ (Step 3).
5. Multiply this square-root result (Step 4) by 2.78.
6. Divide the mean difference (Step 1) by the product of the square root multiplied by 2.78 (Step 5). The final result is the experimental coefficient.

For purposes of illustration, constants describing Distributions 1 and 2 will be used. The measures are listed in tabular form.

Distribution	
1	2
$N = 50$	50
$M = 54.9$	52.5
$\sigma = 19.5$	19.7

Substituting in the formula and following the steps necessary in the computation of the experimental coefficient, the calculations are as follows:

$$\begin{aligned}
 EC &= \frac{54.9 - 52.5}{2.78 \sqrt{\frac{19.5^2}{50} + \frac{19.7^2}{50}}} \\
 &= \frac{2.4}{2.78 \sqrt{\frac{380.25}{50} + \frac{388.09}{50}}} \\
 &= \frac{2.4}{2.78 \sqrt{7.605 + 7.7618}} \\
 &= \frac{2.4}{2.78 \sqrt{15.3668}} = \frac{2.4}{2.78 \times 3.92} = \frac{2.4}{10.8976} = 0.22
 \end{aligned}$$

When EC is 1.0, the difference between the two groups of measures is reliable and significant, the chances being roughly 369 to 1 for significance. The difference 1.0 is such that similar differences between other measures of the same variable will be obtained under similar conditions. Since the difference between Distributions 1 and 2, used in the illustrative problem, is described by an experimental coefficient of 0.22, it is 22 per cent of what it would have to be to indicate a reliable difference between the two distributions.

Readings for Further Study

1. W. A. McCALL, *How to Measure in Education*. Pp. 404-405.
2. W. A. McCALL, *How to Experiment in Education*. Pp. 154-158. New York: The Macmillan Company, 1923.
3. C. W. ODELL, *Statistical Method in Education*. Pp. 355-360.

Often it is desirable to supplement statistical data with a pictorial representation of the frequency distribution. Charts and graphs frequently aid in an understanding of the facts. Certain types of charts that are peculiarly valuable in dealing with measurements situations in classrooms are discussed below.

Graphs.—The two most commonly utilized types of frequency graph are the histogram, or column diagram, and the frequency polygon. The histogram may be constructed quite easily. It is necessary only to locate the mid-point of each class interval on the magnitude scale, or axis of abscissas, and then to erect at each such point an ordinate indicating the frequency in each given class. Lines are then drawn parallel to the base line through the point at the top of each perpendicular to the base line, each parallel line extending in each direction from the point at the height of the ordinate to the limits of the class. These lines are then connected by ordinates drawn from one to the other and perpendicular to the axis of abscissas. The ordinates lie at the class limits. The result is an area completely bounded by the base line, the ordinates at the class limits, and the parallels to the axis of abscissas at the frequency height of each class. Figure 1 represents such a graph for the 50 cases in Distribution 1.

The frequency polygon is plotted somewhat similarly by locating the points at the mid-point of each class interval at the ordinate heights indicating the number of cases in each given class. These points are then connected by straight lines, each point being connected to the next two adjacent points, on either side of the given point. The diagram is then completed by drawing a straight line from the point on the ordinate frequency height in the lowest class interval to the mid-point of the class next below, which has a frequency of 0, and by drawing also a straight line from the ordinate frequency height of the highest class to the mid-point of the class next above, which also has a frequency of 0. Figure 2 represents the frequencies in Distribution 1 in frequency polygon form. It is worth noting that the areas of

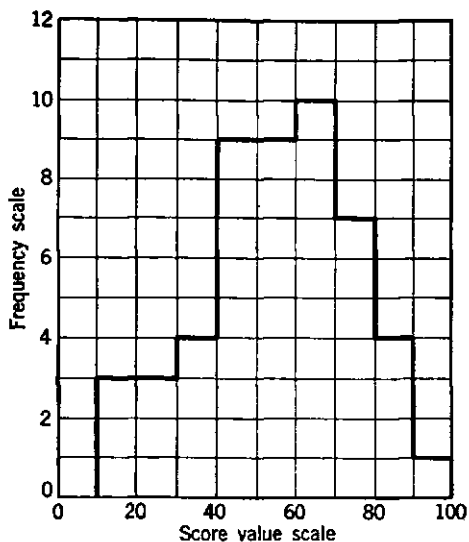


FIG. 1.—Histogram or column diagram, representing the measures in Distribution 1 grouped by class intervals, using 10 as the interval. $N = 50$.

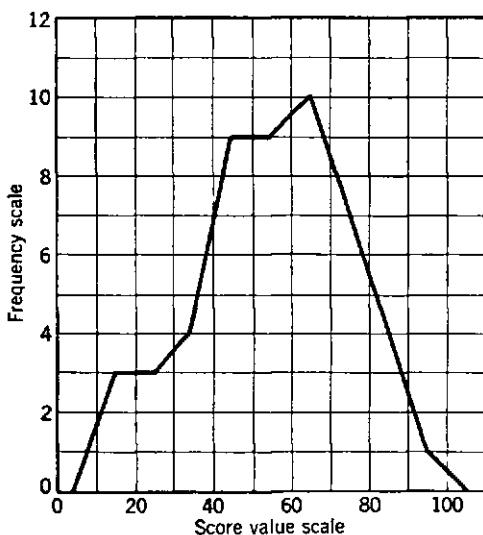


FIG. 2.—Frequency polygon representing the measures in Distribution 1 grouped by class intervals, using 10 as the interval. $N = 50$.

Figs. 1 and 2 are identical, since they represent the same frequency distribution and are drawn on the same scale.

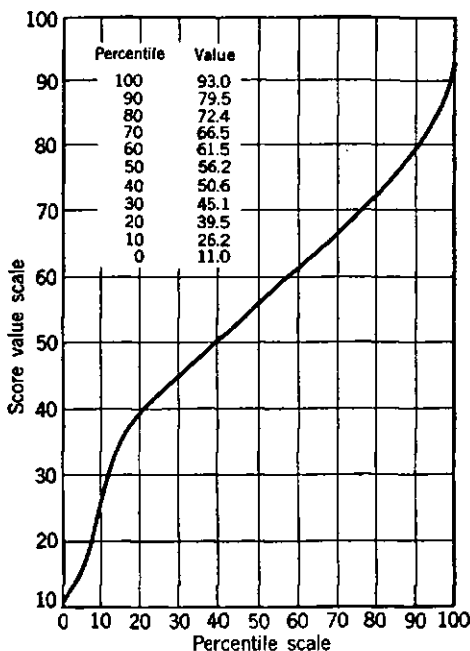


FIG. 3.—Percentile curve or ogive, representing the measures in Distribution 1 grouped by class intervals, using 10 as the interval. $N = 50$.

Figure 3 represents a percentile curve, or ogive, based on Distribution 1. The several percentile values are shown on the chart. The magnitude scale, or scale of values, is placed on the

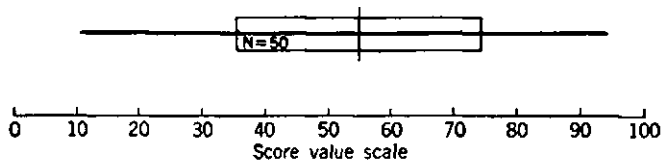


FIG. 4.—Summary chart representing the range (11-93), the mean (54.9), and the standard deviation (19.5) of Distribution 1.

axis of ordinates; and the percentile scale, on the axis of abscissas. All that is necessary is to locate on the ordinate drawn from each given percentile point the numerical value of each given percentile and then to connect these several points by a continuous curved line.

The summary chart, shown in Fig. 4, was developed to show readily the range, the mean, and the standard deviation of a number of distributions of measures that are related each to the others. It shows nothing that may not be given in tabular form, but it does make the facts more readily observed, particularly when many distributions are under consideration. In this chart, the frequency-magnitude scale is located on the axis of abscissas, and an explanatory comment for any given distribution is located on the axis of ordinates. The range is shown by a straight line, parallel to the abscissa axis and extending from a point on the ordinate above the lowest value in the range, as determined on the magnitude scale, to a point on the ordinate above the highest value in the frequency distribution, similarly determined. The mean is located by a short vertical line, intersecting the line graphing the range, located along the ordinate immediately above the mean value, as determined on the frequency-magnitude scale. In graphing the standard deviation, it is necessary to add to and subtract from the numerical value of the mean the numerical value of the standard deviation of the frequency distribution, locating the resulting values on the magnitude scale. At the two ordinate points, two short lines, parallel to the mean line but shorter and intersecting the range line, are drawn. Connecting these two lines, which extend an equal distance on either side of the range line, by straight lines parallel to the range line completes an oblong, which extends an equal distance on either side of the mean ordinate. This oblong indicates the range of $\pm 1\sigma$ on either side of the mean. By making the entry in one corner of the oblong, it is possible to enter the number of cases in the frequency distribution.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 139-141, 148-153.
2. B. D. MUDGETT, *Statistical Tables and Graphs*. Pp. 61-193. Boston: Houghton Mifflin Company, 1930.
3. C. W. ODELL, *Statistical Method in Education*. Pp. 32-65.
4. L. L. THURSTONE, *The Fundamentals of Statistics*. Pp. 9-46.
5. J. E. WERT, *Educational Statistics*. Chap. III.

Comment.—It is impossible in a measurements course to devote too much time to statistical calculation. In the preceding dis-

cussion, only the most important measures have been considered and then only briefly. The student should devote a considerable amount of time to the study of the reference materials listed under each division of the discussion and also to the calculation of typical problems, in order to perfect an ability to do the necessary computation. On the following pages are given a number of practical problems involving educational measurements, taken from actual classroom situations. Practice is afforded in the calculation of the averages, the range, the standard and quartile deviations, the coefficient of correlation, and the experimental coefficient and in the presentation of graphs.

Exercises

1. Determine the range, mean, median, and standard deviation for each score series. Select your interval, and tabulate the scores into a frequency table, using from 7 to 10 classes. Compute *EC*, using your results. Draw one frequency polygon and one histogram, using data in your two frequency tables at least once for graphical purposes.

Score Series 1:

93	89	73	21	47
14	19	81	14	44
62	92	87	48	17
39	86	83	33	52
24	23	37	16	56

Score Series 2:

100	49	73	91	28
43	33	66	75	55
99	21	55	54	66
47	64	79	39	77
56	44	88	51	43

2. Using the life ages and the Terman total scores given below, obtain mental age and I.Q. for each of the 25 pupils in group A and for each of the 25 pupils in group B. Then compute for each distribution of I.Q.'s the following measures: the range, the mean, the median, the standard deviation. Using data for both series, compute *EC*. Present your data graphically, using the histogram and the horizontal line, the vertical line, and the open oblong technique (for a summary of the range, the mean, and the standard deviation). Determine the coefficient of correlation for the paired life ages and Terman scores.

Terman total score	Stanford Binet mental age	Terman total score	Stanford Binet mental age
170	17-7	110	14-10
165	17-5	105	14-7
160	17-2	100	14-4
155	16-11	95	14-1
150	16-8	90	13-11
145	16-5	85	13-8
140	16-3	80	13-5
135	16-0	75	13-2
130	15-9	70	12-11
125	15-6	65	12-9
120	15-3	60	12-6
115	15-1	55	12-3

Group A		Group B	
Life ages	Terman scores	Life ages	Terman scores
16-2	166	15-10	93
13-11	151	14-11	74
14-11	93	14-9	122
12-8	137	11-11	142
15-2	117	12-11	70
15-3	86	14-7	131
14-2	98	13-11	116
11-11	128	14-1	137
12-3	117	14-3	72
16-1	93	14-10	110
14-4	106	14-7	106
13-10	97	15-1	95
12-9	111	14-3	141
13-4	93	16-1	98
15-4	128	15-7	118
14-8	121	14-8	116
14-1	107	13-11	104
13-9	97	14-1	139
12-11	99	14-7	108
15-6	125	14-4	100
15-3	75	15-10	116
13-7	101	13-11	97
14-8	99	12-10	99
13-7	112	11-11	108
12-10	151	13-9	142

3. Using the I.Q.'s below, group the pupils into four sections of 25 each. In Group 1, place the 25 highest I.Q.'s; in Group 2, the 25 next highest I.Q.'s; in Group 3, the 25 next lowest I.Q.'s; and in Group 4, the 25 lowest I.Q.'s. Determine for each such group, using achievement scores *only*, the range, mean, and standard deviation in each group. Compute the experimental coefficient for each group with each other group. Determine the coefficient of correlation for the paired I.Q.s and achievement scores.

I.Q. (Terman)	Achievement score	I.Q. (Terman)	Achievement score
105	50	85	13
89	26	85	24
121	38	113	28
100	46	120	47
103	46	88	19
106	53	86	36
102	35	106	47
110	55	89	41
95	14	98	22
105	31	84	41
108	44	117	56
98	30	94	27
97	44	86	47
105	47	107	25
80	24	110	48
113	64	107	54
98	39	94	37
108	48	94	41
101	35	98	18
100	43	87	23
117	23	95	38
87	27	103	56
106	24	113	34
125	55	106	49
111	51	125	51
116	36	79	31
111	15	111	10
98	47	105	14
87	28	116	44
93	23	89	18
87	28	101	42
100	38	85	25
104	27	103	53
103	31	83	17
98	27	107	22
112	36	129	32
108	45	107	49
107	25	113	34
98	28	126	57
92	32	109	52
107	45	96	45
77	35	107	26
100	21	116	46
86	23	94	34
85	25	104	44
97	23	100	56
93	36	94	28
99	51	101	38
92	21	98	46
71	26	94	45

4. Translate into age scores and into grade placement scores, for the pupils in Groups A and B, the pupil scores below, using the table of norms given below. Compute (1) the range, (2) the mean, and (3) the standard deviation for each of the four (derived score) frequency distributions. Compute also (4) the two *ECs*, one for age scores and the other for grade placement scores. Present the data graphically, using (a) the histogram, and (b) the horizontal line-vertical line-open oblong technique (for showing the range, the mean, and the standard deviation).

Grade placement	Total score	Age, months	Grade placement	Total score	Age, months
0.6	0	71	2.0	25-26	88
0.7	1	72	2.1	27-28	89
0.8	2	73	2.2	29-30	90
0.9	3-4	74	2.3	31-32	92
1.0	5-6	75	2.4	33-34	93
1.1	7-8	76	2.5	35-36	94
1.2	9-10	77	2.6	37-38	96
1.3	11-12	79	2.7	39-40	97
1.4	13-14	80	2.8	41-42	98
1.5	15-16	81	2.9	43-44	100
1.6	17-18	82	3.0	45-46	101
1.7	19-20	84	3.1	47-48	102
1.8	21-22	85	3.2	49	104
1.9	23-24	87	3.3	50	105

Pupil scores: (Grade 1A, first of semester test findings.)

Group A:					Group B:				
9	8	22	11	11	5	19	14	10	24
21	7	20	8	8	11	6	21	8	27
14	4	13	13	6	10	14	11	13	19
11	13	8	7	17	21	12	17	7	14
11	20	14	8	31	3	10	8	18	9

5. Pupils in grade 5B of an elementary school were given the New Stanford Achievement Test in January, prior to promotion. Use the scores below to compute coefficients of correlation for the paired series of scores. Compute means and standard deviations from the data in the scattergraphs.

NEW STANFORD ACHIEVEMENT TEST SCORES, 25 PUPILS

Para- graph mean- ing	Word mean- ing	Total read- ing	Spell- ing	Lan- guage usage	Litera- ture	His- tory and civics	Geog- raphy	Physi- ology and hy- giene	Arithmetic		
									Rea- son- ing	Com- puta- tion	Total
72	75	147	68	91	75	72	75	69	75	67	142
77	70	147	56	88	75	64	69	72	69	62	131
80	62	142	65	92	47	68	63	83	83	61	144
60	63	123	62	62	69	62	58	63	58	67	125
67	39	106	44	20	50	57	50	50	65	59	124
93	99	192	82	95	86	88	98	90	88	65	153
96	93	189	61	102	93	89	99	78	65	58	123
100	91	191	72	90	86	83	81	76	78	65	143
92	84	176	53	97	88	81	75	84	72	59	131
92	81	173	60	104	76	78	73	82	62	59	121
88	85	173	46	85	81	80	81	77	75	62	137
87	81	168	68	90	73	72	74	74	72	62	134
88	80	168	65	83	79	74	75	74	62	59	121
84	84	168	58	88	79	47	67	85	62	59	121
92	78	170	39	57	93	80	58	69	83	64	147
69	74	143	63	74	75	72	76	76	58	56	124
66	66	132	50	88	69	66	71	65	72	64	136
87	81	168	60	62	73	54	57	69	62	59	121
74	70	144	44	81	50	74	61	58	58	64	122
70	69	139	48	46	62	72	64	69	65	62	127
74	66	140	42	48	90	35	70	71	37	58	95
67	74	141	66	66	20	43	64	73	48	61	109
59	59	118	50	20	67	54	61	73	58	59	117
44	65	109	25	44	62	66	56	53	37	56	93
50	46	96	33	36	50	54	62	23	46	51	99

6. The Progressive Achievement Test, Form A, was administered to pupils in grade 6A of an elementary school in February, at the beginning of the semester. Use the grade placement scores for the 25 pupils in the several subjects to determine (1) the median, (2) the mean, (3) the standard deviation, and (4) the percentage of pupils in the group in each subject that reached and exceeded the norm (6.5). (NOTE.—the paired series of scores may also be used to find coefficients of correlation, if desired.)

PROGRESSIVE ACHIEVEMENT TEST SCORES, 25 PUPILS

Reading vocabulary	Reading comprehension	Arithmetic reasoning	Arithmetic fundamentals	Language	Totals
8.9	8.6	8.2	6.5	7.6	7.6
9+	8.2	7.7	5.7	7.5	7.5
8.9	8.0	8.8	5.8	7.3	7.3
8.3	7.9	7.1	5.6	8.0	7.2
5.6	7.0	6.8	6.3	7.5	6.9
6.9	6.6	8.2	6.4	6.2	6.8
6.8	7.2	7.0	5.3	6.8	6.5
6.9	7.5	6.2	5.2	7.5	6.5
7.3	7.2	6.0	5.3	7.5	6.5
6.6	7.7	7.0	5.2	6.5	6.4
7.3	6.9	6.2	5.3	6.7	6.4
6.1	7.0	7.4	5.8	5.8	6.3
8.8	7.4	8.0	4.8	5.3	6.2
6.1	6.5	6.2	5.1	7.5	6.2
5.7	6.5	5.6	4.9	7.6	6.1
6.0	5.3	5.7	5.3	7.0	6.0
6.0	5.9	7.4	5.4	5.6	5.9
6.1	5.5	6.6	5.7	6.2	5.9
6.6	6.6	6.3	5.2	6.0	5.9
6.5	6.2	6.6	5.2	5.3	5.8
5.3	6.8	7.5	5.0	5.5	5.6
4.5	5.4	7.1	5.6	5.2	5.5
5.9	5.2	4.5	5.2	4.4	5.1
5.6	5.2	4.2	5.1	5.4	5.1
7.0	4.6	4.5	4.1	4.3	4.8

7. Pupils in grade 6A in an elementary school studied for six weeks an activity unit dealing with the exploration of North America. The teacher used two objective tests, each of 40 items, to measure pupil achievement. Test I was used as the pretest and as one-half the post test; Test II, as the second half of the post test. Using the scores for 25 pupils given below, prepare the measures needed in summarizing achievement gains and in justifying the validity of the gains: (1) means and standard deviations, (2) medians and percentages of pupils on the first half of the post test reaching and exceeding the median for the pretest, (3) the experimental coefficient for the pretest and the first half of the post test, and (4) the coefficient of correlation involving the paired scores for the two halves of the post test.

Test I pretest	Test I $\frac{1}{2}$ post test	Test II $\frac{1}{2}$ post test
17	37	39
12	36	39
10	33	38
8	37	37
16	40	39
11	38	39
12	37	37
14	32	36
10	38	36
19	39	39
13	35	36
5	26	33
10	38	39
19	40	40
5	35	39
11	37	39
8	32	31
9	31	34
14	37	39
9	35	39
13	38	37
14	39	38
15	39	40
10	35	32
9	39	36

CHAPTER IV

THE USE AND INTERPRETATION OF STATISTICAL MEASURES

It is not enough for the classroom teacher who is a user of statistics to be able to compute the statistical constants described in Chapter III and to know the nature of each such statistical measure and its advantages and limitations. She must also be able to relate statistical theory and obtained statistical data and the classroom situation in which the data were obtained. Statistics, properly used, yield information as to *what* a situation is, but they never tell *why* it is as it is or *how* it came to be as it is. The classroom teacher must be able to analyze statistical evidence in order to understand a given situation, in so far as statistical data describe it, and she must be able also to relate the results to their causes and to trace the trend of development that led to the findings obtained. These things are necessary if data yielded by measuring instruments are to be used to check the proper functioning of an instructional program or as a basis for planning modifications in instructional methods and procedures leading to increased teaching efficiency.

The purpose of this chapter is to present the fundamentals needed by a classroom teacher in thinking about statistical results.¹ The discussion will deal only with those measures which have been presented in Chapter III, and the order of presentation will parallel that of Chapter III.

Measures of Central Tendency. *The Mean.*—Many classroom situations involve comparisons of summaries of statistical series. For example, it is possible to study gains due to instruction by studying the achievement of pupils as evidenced by test scores yielded by the same test or by duplicate and equivalent forms of a test, administered to pupils before and after a period of instruc-

¹ The inquiring student will find in the references for further study a wealth of information bearing on this subject which is excluded from this discussion because of the necessary brevity of the treatment.

tion. Similarly, it is possible to compare pupil achievement from year to year and pupil achievement involving students in two or more class groups at a given time by using a single achievement test. This is done frequently when the effects of different methods of teaching subject matter are being considered. Comparisons involving the use of statistical data often are not only desirable but necessary.

It is necessary to have a reference point, or point of departure for thought, when one is making comparisons of statistical series. The best reference point is an average, since an average is the most typical value in a collection of measures. Ordinarily, the mean is the most frequently used measure of central tendency.

The mean is a computed average, based on the actual values of all the items included in the aggregate of measures. The value of the mean is the value that would be found if all of the measures included in the collection of data had the same numerical value, and the number of cases remained unchanged. The sum of the deviations of the individual measures from the mean, determined algebraically, is zero. When groups being considered are small, the mean may not be the best average to use for purposes of comparison, since it is affected materially by extreme deviations from the central tendency. One extreme deviation value offsets several values near the central tendency in determining the numerical value of the mean. Despite this limitation affecting its usefulness, the mean should be computed always, and it should be employed invariably as the measure of the central tendency of a series of measures, since it is itself an actual score value, unless there is some very definite reason for the employment of the median.

Whenever the statistical analysis involves the calculation of coefficients of correlation, the mean should be reported. It may be obtained readily from the calculations required for determining the value of the coefficient of correlation, particularly when the scattergraph technique is utilized. Similarly, the mean is necessary when groups are compared by means of critical ratios.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 42-45, 187-195.
2. R. E. CHADDOCK, *Principles and Methods of Statistics*. Pp. 103-105. Boston: Houghton Mifflin Company, 1925.

3. H. E. GARRETT, *Statistics in Psychology and Education*. P. 29.
4. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 83-85.
5. C. W. ODELL, *Statistical Method in Education*. Pp. 98-100.

The Median.—The median, unlike the mean, is not based on the actual values of the items included in an aggregate of measures. It is a point on the numerical scale of values; it is not an actual score value in the same sense as is the mean. The median serves merely to divide the distribution of items exactly into two equal parts, one-half of the cases lying above it; and the other half,

Distribution	
A	B
66	66
60	60
60	60
60	60
54	60
6	54
<u>6)306</u>	<u>6)360</u>
$M = 51$	60
$Md = 60$	60

below it. It should be used, alone or with the mean, when it gives or helps to give a good summary of a frequency aggregate. The median is particularly valuable for this purpose when it differs in value from the mean. For example, in Distribution A, the median is 60; the mean, 51. It is obvious from an inspection of the series of measures in this instance that the value of the mean is seriously decreased by the presence of one measure removed in value far below the remaining values in the frequency aggregate. In this instance, the median is the better measure of central tendency, since it is more typical of the whole series than is the mean. In Distribution B, however, the mean and the median values are identical, both being 60. In this instance, the median yields no information not given also by the mean, and the mean should be used as the measure of central tendency.

The median should be computed not only when it is desired to avoid the influence of extreme or erroneous values in the frequency distribution or when the exact values of all the items in the original frequency aggregate are not known¹ but also when-

¹ It is possible to determine a median when only a part of a total frequency aggregate is known, e.g., the middle 50 per cent or the middle 80 per cent of

ever it is desired to compute percentile values other than the median to obtain a complete characterization of the structure of the frequency distribution. When standard test scores are being considered, it is almost always necessary to calculate the value of the median, since the majority of test norms are expressed as medians. It is not possible to compare satisfactorily the mean score, expressing average achievement for a given group of pupils, with the norm of a standardized test which is a median.

Because it is a percentile measure, the median lends itself readily to mass comparisons involving two or more frequency aggregates. Thus, for example, a comparison may be made between Distributions *A* and *B*, by saying that 83.33 per cent of the measures in Distribution *B* reach or exceed the median value 60 of Distribution *A*.

In ordinary practice, it seems best to compute the values of both the mean and the median and to report both averages in summarizing any given frequency distribution. Both have specific values, and reporting both makes all of the information available. This practice also permits the use of either or of both measures of central tendency in the study of the situation under consideration.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 45-47.
2. R. E. CHADDOCK, *Principles and Methods of Statistics*. Pp. 122-124.
3. C. C. PETERS and W. R. VAN VOORHIS, *Statistical Procedures and Their Mathematical Bases*. Pp. 47-48. State College, Pennsylvania: School of Education, The Pennsylvania State College, 1935.
4. G. U. YULE, *An Introduction to the Theory of Statistics*. Pp. 118-120.

Measures of Dispersion. *The Range.*—The range is the distance on the linear scale of values from the lowest to the highest measure. It shows, therefore, the total numerical distance included in the frequency aggregate. It is obtained readily, but it is unreliable, being influenced by extreme departures from the central tendency; therefore, it is the least useful of the several measures of variability. Ordinarily, however, it should be obtained, and particularly this is true when a knowledge of the total spread of the values in the frequency series is all that is

the cases. The median in such instances can be determined by finding the mid-point of the available values.

wanted or when the values are too few in number or too scattered to justify the calculation of any other measure of dispersion.

The Standard Deviation.—The standard deviation includes in its range on either side of the mean, with which it is associated by definition, the middle 68.27 per cent of the cases included in the frequency aggregate, when the distribution is symmetrical and the items included are distributed normally. For convenience, one may think of the standard deviation as including about the middle two-thirds of the cases in any given distribution.

It is not enough to know the average measures for frequency distributions that are being compared; information also must be given as to their homogeneity. The standard deviation serves very well for this purpose; a large standard deviation reveals the presence in the frequency aggregate of many measures departing far from the central tendency. A standard deviation of large value, then, reveals a lack of normal consistency in the frequency series. It indicates that the cases included in the aggregate tend to be heterogeneous rather than homogeneous. The smaller the value of the standard deviation the greater is the concentration of the measures in the frequency series about the central tendency.

Class	Distribution		
	X	Y	Z
90 —	5	20	10
80 —	10	10	10
70 —	10	10	15
60 —	15	5	10
50 —	20	10	10
40 —	15	5	10
30 —	10	10	15
20 —	10	10	10
10 —	5	20	10
<i>N</i> =	100	100	100
<i>Md</i> =	54.5	54.5	54.5
<i>M</i> =	54.5	54.5	54.5
σ =	21.2	30.2	25.3

In Distribution X, Y, and Z, the values of the mean and the median have been made to be identical. This might lead an

unwary observer to think of these three distributions as being similar. Such is not the case, as may be observed by studying the three frequency arrays. Such a study is unnecessary, however, since the values of the three standard deviations show clearly that the three distributions are not similar, even though the several measures of central tendency are identical in value.

The method of computing the standard deviation insures the obtaining of a refined, accurate, and statistically significant measure of dispersion. The standard deviation is mathematically not difficult to obtain. It may be secured quite readily from the scattergraph computations involved in finding the coefficient of correlation, simply by multiplying the obtained standard error value of the given frequency series by the value of its interval. It is a convenient measure to use in making comparisons of distributions having unequal or unknown zero points and unit scale values, particularly when the analysis involves percentile curves based on such differing values. In the same way, the standard deviation is of value in determining equivalent score values for score series having differing zero points and unit scale values. All that is necessary is to find the distance of given scores from the mean of the distribution, in terms of the standard deviation value.

Distribution	
I	II
$N = 50$	50
$M = 60$	82
$\sigma = 10$	15

For example, a value of 54 in Distribution I would lie 0.6σ below the mean of the distribution 60. An equivalent score in Distribution II would have a value of 73, 0.6σ below the mean. Similarly, a score value of 70 in Distribution I would be equivalent to a score value of 97 in Distribution II. In both instances, the equivalent scores would lie the same standard deviation distance from the means of their respective distributions and in the same direction from the means.

When comparisons are made of two or more distributions, involving the use of the critical ratio, it is necessary first to compute the standard deviations of the frequency aggregates to be compared. The standard deviation also has been used in marking procedures. Using the mean as a point of departure and

computing the values of $\frac{1}{2}\sigma$ and of $1\frac{1}{2}\sigma$, it is possible to locate failing marks below $-1\frac{1}{2}\sigma$, poor (but passing) marks between $-1\frac{1}{2}\sigma$ and $-\frac{1}{2}\sigma$, average marks between $-\frac{1}{2}\sigma$ and $+\frac{1}{2}\sigma$, good marks between $+\frac{1}{2}\sigma$ and $+1\frac{1}{2}\sigma$, and excellent marks above $+1\frac{1}{2}\sigma$. This yields almost exactly the familiar 7-24-38-24-7 percentage distribution of marks.

The real utility of the standard deviation is not in these auxiliary employments but rather in describing the character of frequency aggregates, by indicating their homogeneity, *i.e.*, the tendency of items to cluster close about or to depart from the central tendency. The standard deviation should be used always as the measure of variability, unless there is some very definite reason for using some other measure, and even then probably it should be included among the measures used in summarizing the facts of a frequency series.

Readings for Further Study

1. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 63-65, 195-203.
2. R. W. BURGESS, *Introduction to the Mathematics of Statistics*. Pp. 147-149.
3. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 113-116.
4. H. JEROME, *Statistical Method*. Pp. 250-255. New York: Harper & Brothers, 1924.

The Quartile Deviation.—This measure is associated with the median in much the same way that the standard deviation is associated with the mean. The quartile deviation demands the computation of percentile 25 (quartile 1) and of percentile 75 (quartile 3). The quartile deviation is the value of one-half the distance between these two percentile points. When the frequency distribution is symmetrical and the items which it includes are distributed normally, percentiles 25 and 75 are equidistant from the median. Very frequently, however, distributions are skewed, and the two percentile points are not located equidistant from the median. In such an instance, the quartile deviation may yield misleading information concerning the actual concentration of the frequencies about the central tendency.

Ordinarily, the quartile deviation should be used to describe variability when only the degree of concentration about the median is sought and when only the median is given as a measure

of central tendency; when the numbers of cases are relatively few and are widely scattered; when there are several cases far removed from the central tendency (and in this instance both the standard deviation and the quartile deviation should be reported); and when norms for some standardized test involved in the problem under consideration include percentiles other than the median. In the latter instance, obtaining the quartile deviation permits comparisons with test norms at three points rather than at one point only, the median.

Examples of Uses of Measures of Central Tendency and of Dispersion and of Graphs.—It is difficult to make clear the uses and the significance of the measures of central tendency and of dispersion that have been discussed; and examples, together with interpretative comment, are necessary if this is to be achieved. Certain typical materials have been chosen for presentation.

The first of these excerpts deals with differences in citizenship, as determined by rating scales yielding point scores, among pupils of differing levels of mental ability and of introversion-extroversion. The data are presented in Table I.

TABLE I.—SUMMARIES OF CITIZENSHIP POINT RATINGS FOR FOUR GROUPS OF PUPILS GROUPED ACCORDING TO DIFFERING MENTAL ABILITY AND PERSONALITY LEVELS¹

$N = 122$
Terman Mental Age
Below the Md $Md = 13 - 4$ Above the Md

MacNitt extroversion- introversion score $Md = 116.2$	Above the Md	$N = 32$ Range = 29 - 116 $M = 80.4$ $\sigma = 19.8$	$N = 29$ Range = 37 - 130 $M = 93.1$ $\sigma = 23.6$
	Below the Md	$N = 29$ Range = 42 - 115 $M = 80.4$ $\sigma = 20.4$	$N = 32$ Range = 42 - 130 $M = 88.3$ $\sigma = 21.3$

¹ M. E. BROOM and JOSEPHINE D. BRAMKAMP, A Study of Factors Contributing to School Citizenship, *Journal of Applied Psychology*, 17: 466 (1933).

Certain facts may be deduced from Table I. It would appear that, on the whole, pupils above the average are better citizens

in school than are pupils below the average in mental ability. The differences in citizenship, on the average, are greater for pupil groups of differing mental ability levels than for pupils of differing personality characteristics. Among pupils below the average in mental ability, no citizenship average differences result among pupils of differing personality characteristics. Homogeneity in citizenship, evidenced by the standard deviation, is greater among pupils below than among those above the average in mental ability. In so far as the data in Table I are concerned, there seems to be no relationship between personality and homogeneity in citizenship; the results show neither personality group favored by small standard deviations both above and below the median mental age.

The second brief study deals with grade point ratios *GPR*, the mean percentage grade earned for all subjects taken during four years of high-school study; with scores yielded by a teacher-made objective test of general business information; and with intelligence quotients yielded by the Henmon-Nelson Test of Mental Ability. Table II includes summaries of the three series of measures, together with percentiles for each series.

TABLE II.—MEASURES SUMMARIZING AND DESCRIBING THREE SPECIFIED SERIES OF MEASURES¹
N = 270

Measure	<i>GPR</i>	Business backgrounds test scores	I.Q.
Range	77.3 to 93.2	32 to 143	66 to 128
<i>M</i>	83.9	77.8	108.5
σ	3.6	22.5	11.2
<i>Per</i> ₁₀₀	93.2	143.0	128.0
<i>Per</i> ₉₀	89.0	109.1	123.3
<i>Per</i> ₈₀	87.4	99.2	118.3
<i>Per</i> ₇₀	86.0	89.8	115.3
<i>Per</i> ₆₀	84.6	83.0	112.4
<i>Per</i> ₄₀ (<i>Md</i>)	83.3	76.7	109.4
<i>Per</i> ₄₀	82.4	70.4	106.1
<i>Per</i> ₂₀	81.5	62.0	102.8
<i>Per</i> ₂₀	80.7	56.3	99.5
<i>Per</i> ₁₀	79.8	48.6	93.5
<i>Per</i> ₀	77.3	32.0	66.0

¹ The data were furnished by Mathilde Hardaway, Austin High School, El Paso, Tex.

Since these three variables are known to utilize differing unit values, and since the three series are not known to have a uniform zero point, it becomes impossible to compare them directly in order to determine whether or not the characteristics of the three distributions are similar. There is a way by means of which this may be done, however; all that is necessary is to find the standard deviation distance of each percentile point in a given distribution from its mean. The data for the three distributions of measures being considered are given in Table III.

TABLE III.—STANDARD DEVIATION DISTANCES OF CERTAIN PERCENTILE POINTS FROM THE MEANS OF THEIR RESPECTIVE DISTRIBUTIONS

Percentile	<i>GPR</i>	Business backgrounds Test scores	I.Q.
100	+2.58	+2.90	+1.74
90	+1.42	+1.39	+1.32
80	+0.97	+0.95	+0.88
70	+0.58	+0.53	+0.61
60	+0.19	+0.23	+0.35
50 (<i>Md</i>)	-0.17	-0.05	+0.08
40	-0.42	-0.33	-0.21
30	-0.67	-0.70	-0.51
20	-0.89	-0.96	-0.80
10	-1.14	-1.30	-1.34
0	-1.83	-2.04	-3.79

The data presented in Table III have been assembled graphically in Fig. 5. This chart yields no information not included in Table III, but it does present it in more readily comprehensible fashion. This is the function of graphs, to present data in easily comprehensible form. The data in Table III and Fig. 5 show clearly that the three distributions have similar characteristics, excepting the extremes of the distribution of intelligence quotients. It should be noted that in this technique, means and standard deviations are used regardless of their numerical value. The mean in each instance is used as a point of departure; the standard deviation, as a unit of measure.

The third presentation is more inclusive than either of the preceding ones. It is concerned with a summary of reading achievement in 14 half-grade groups in a public-school system. The

data were obtained in 22 schools during the second semesters of 1936-1937 and of 1937-1938.

Certain facts stand out plainly in this tabulation. The median tends to exceed the mean in value in the high-school grade levels, indicating that the mean is influenced materially by extremely low scores earned by certain pupils. In the elementary-school

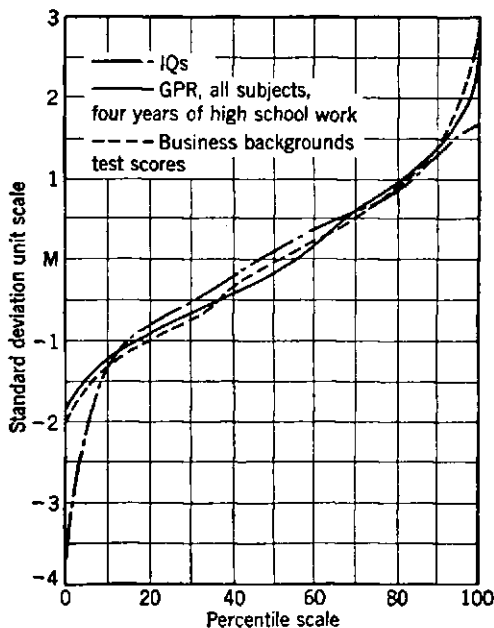


FIG. 5.—Percentile chart for three given variables, measured in differing unit values. The mean and the standard deviation are taken as uniform quantities regardless of their numerical value. $N = 270$.

grades, the reverse situation is found. Homogeneity is quite uniform among the high-school groups, seven of the eight standard deviation values being but a fraction of a score point removed from 14.0. Similarly, homogeneity tends to be uniform among the elementary groups although less than among high-school groups. The standard deviations for four of the elementary-school groups are within a fraction of a score point of 16.0, and the two remaining standard deviations vary from 16.0 by about 10 per cent of this value. It is worth noting that the averages show a reasonably steady progression in improved reading achievement, although there are two inversions: (1) grade H 10 has a higher mean score and a lower median score than grade

L 11, and (2) grade H 9 has a higher mean and median than the next higher three half grades. Gains from grade to grade in reading achievement are smaller among the higher than among the lower grades. Taking all the summary measures into

TABLE IV.—SUMMARY OF READING SCORES FOR 14 HALF GRADES IN A PUBLIC-SCHOOL SYSTEM¹

$N = 7,873$

Grade	N	Range		Md	M	σ
		From	To			
H 12	399	41	99	83.3	81.3	12.5
L 12	203	39	100	82.0	79.7	14.4
H 11	410	31	97	81.1	78.2	13.8
L 11	327	33	97	78.1	75.8	13.5
H 10	499	4	98	79.1	76.0	14.6
L 10	395	0	98	73.9	75.1	14.6
H 9	618	20	100	80.2	77.0	13.9
L 9	497	26	100	71.8	71.3	14.6
H 8	688	25	96	68.6	68.1	15.9
L 8	604	16	96	60.9	61.1	16.7
H 7	806	10	99	54.9	55.7	15.8
L 7	689	11	90	44.6	46.5	15.6
H 6	933	4	90	40.1	42.1	17.4
L 6	785	2	87	28.4	30.8	14.4

¹ These data were compiled by the writer for use by the Department of Instruction of the El Paso, Tex., Public Schools.

account, it is safe to say that this test is more discriminative among pupils in lower than among those in higher grades. It appears also that this test does not provide sufficient opportunity for pupils in advanced grades to demonstrate superior ability; the maximum possible score is too often achieved by pupils in these grades, and the distance between the upper limit of $+1\sigma$ and the maximum possible score is too narrow.

Figure 6 presents the data from Table IV in graphical form. The chart gives the range, the mean, the median (as a norm), and

the standard deviation for each score series. As with Fig. 5, this chart does not yield any information not given in Table IV, but it does present it in much more comprehensible form. It would be impossible to get a clear-cut concept of the achievement situation in the 14 half grades from a casual inspection of Table

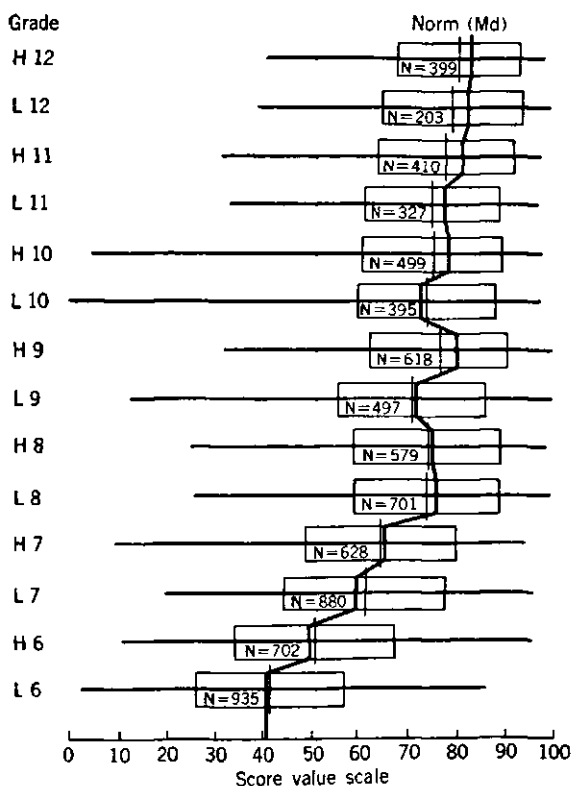


FIG. 6.—Chart presenting summary data on silent reading achievement in fourteen half grades in elementary and high schools in the El Paso Public Schools, 1937-1938.

IV, but such a casual inspection of Fig. 6 will yield a reasonably comprehensive knowledge of the facts.

The effort in presenting these three examples has been to indicate correct tabular form and graphical form for presenting data and to show certain uses for the different statistical constants. In addition, an attempt has been made to show typical school situations in which statistics are of value and to show also how one should think with reference to statistical data. The

exercises that follow this chapter will include similar data, also taken from actual school situations and providing the student with opportunity to develop skill in interpreting statistical data.

Relationship. *The Coefficient of Correlation.*—The interpretation of the meaning of an obtained coefficient of correlation is not at all easy. Frequently before the meaning of the coefficient can be determined, certain additional computations must be made, and the meaning of the resulting values considered. It is unnecessary for the classroom teacher using statistics for practical instructional purposes to apply all of the measures in determining the significance of an obtained coefficient that would be applied by a research worker interested in analyzing minutely a complex problem. Certain general conclusions may be formulated which may serve to guide the classroom teacher as to the meaning of coefficients of correlation of varying values.

A coefficient of correlation may be thought of as having two types of significance. In the first place, an obtained coefficient measures relationship between the two specific groups of paired measures. In the second place, the coefficient may be thought of as indicating the relationship that exists, in general, between paired scores for the two variables under consideration. Always, the coefficient of correlation expresses correspondence between paired variables that demonstrate a rectilinear relationship.

When the paired series of measures tend to vary together, the high scores in one series corresponding in pairs to the high scores of the other, the relationship and the value of the coefficient are positive. When the paired series of measures tend to vary inversely, the high scores in one series being paired with low scores in the second series, the relationship and the value of the coefficient are negative.

It is possible to set up tentatively a classification of values for coefficients of correlation, with general statements as to the meaning of values at each level of value. Such a general guide follows: r of a value between 0.00 and ± 0.20 denotes negligible relationship; r of a value from ± 0.20 to ± 0.40 denotes low relationship; r of a value from ± 0.40 to ± 0.60 denotes definite but relatively low relationship; r of a value from ± 0.60 to ± 0.80 denotes marked to substantial relationship; and r of a value from ± 0.80 to ± 1.00 denotes high relationship to perfect positive or negative correspondence.

It is now possible to relate these value levels for coefficients of correlation to reliability and to validity. When the reliability coefficients for a test are computed for paired series including pupil scores obtained in several adjacent grade levels, they should not fall below 0.90. Tests that earn reliability coefficients under such conditions may be considered as having a satisfactory degree of consistency for group or for individual measurement. Reliability coefficients for pupil scores on one grade level only may fall as low as 0.80, and tests having coefficients measuring consistency above 0.80 under these conditions may be thought of as being satisfactory for group measurement but only fairly satisfactory for individual measurement.

When scores yielded by two tests purporting to measure the same thing are correlated, the resulting coefficient indicates the degree of validity or the extent to which the two tests validate each other. The minimum level for validity coefficients obtained under these conditions should be $+0.60$, and a desirable level is $+0.70$. Table V presents data showing how five handwriting scales may be used, each to validate the others, since each has claim to validity by virtue of the procedures used in standardizing each scale.

TABLE V.—MEASURES OF THE VALIDITY OF FIVE HANDWRITING SCALES¹
N = 200

Scale	Leamer	Starch	Thorndike	Zaner
Ayres.....	0.65	0.63	0.67	0.73
Leamer.....		0.73	0.64	0.88
Starch.....			0.65	0.76
Thorndike.....				0.81

NOTE.—The probable errors vary from 0.01 to 0.03.

¹ H. TENWOLDE, A Study of the Penmanship of Pupils in an Elementary School System, p. 36, unpublished Master's thesis, University of Southern California, 1932.

In Table V, all of the scales, always using four scales as criterion instruments, demonstrate a satisfactory degree of validity. It is possible to use this technique in determining the validity of a test, validity of which is unknown, simply by using scores yielded by a valid test as the criterion.

There are occasions when high relationship is *not* desired as evidence of validity. Such a situation arises when several instru-

ments are used in measuring achievement in related fields or in differing phases of a field. Such a situation is presented in Table VI.

TABLE VI.—INTERCORRELATIONS OF FOUR UNIT SUMMARY TESTS IN
GENERAL SCIENCE¹
N = 100

Unit	II	III	IV
I	0.38	0.54	0.41
II	0.63	0.19
III	0.24

¹J. S. CARROLL, *The Bases, Construction, and Statistical Analysis of a Series of Objective Unit Summary Tests in General Science*, p. 30, unpublished Master's thesis, University of Southern California, 1932.

Unit I dealt with astronomy; Unit II, with construction methods; Unit III, with machines; and Unit IV, with electricity. Each of the tests included 20 true-false items, 10 four-answer multiple-response items, 5 two-item completion exercises and 5 one-answer completion items, and 20 identification items. Of necessity, there was some overlapping of the content of Units II and III. The coefficients in Table VI clearly show that the units, with the exception noted, included different content, if the test results reflect accurately what was taught. The relationships in Table VI are all positive, indicating that the four tests are measuring some phase or phases of the same thing, and fairly low in value. Thus, each test tends to stand alone as an individual measuring instrument.

This treatment is necessarily meager, and it is not intended to give the research worker the knowledge that he should have concerning the interpretation of the coefficient of correlation. Rather, the effort has been made to provide the minimum of information that will be of value to the classroom teacher using statistics. The inquiring student will find full information in the references for further reading.

Readings for Further Study

1. A. S. BARR, *The Coefficient of Correlation*, *Journal of Educational Research*, 23: 55-60 (1931).
2. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 86-94, 203-211.

3. W. BROWN and G. H. THOMSON, *The Essentials of Mental Measurement*. Pp. 97-106. New York: The Macmillan Company, 1921.
4. P. H. FURFEY and J. F. DALY, Product-moment Correlation as a Research Technique in Education, *Journal of Educational Psychology*, 26: 206-211 (March, 1935).
5. P. H. FURFEY and J. F. DALY, The Interpretation of the Product-moment Correlation Coefficient, *Catholic University of America Educational Research Monographs* 4, Vol. VIII. Washington: The Catholic University Press, 1934.
6. H. E. GARRETT, *Statistics in Education and Psychology*. Pp. 341-356.
7. K. J. HOLZINGER, *Statistical Methods for Students in Education*. Pp. 163-167, 172-173.
8. W. S. MONROE, *An Introduction to the Theory of Educational Measurement*. Pp. 344-358.
9. W. S. MONROE and D. B. STUTT, The Interpretation of the Coefficient of Correlation, *Journal of Experimental Education*, 1: 186-203 (March, 1933).
10. P. H. NYGAARD, Interpretation of Correlation on the Basis of Common Elements, *Journal of Educational Psychology*, 23: 578-585 (1932).
11. C. W. ODELL, *Statistical Method in Education*. Pp. 188-208.
12. C. W. ODELL, The Interpretation of the Probable Error and the Coefficient of Correlation. *Bureau of Educational Research Bulletin* 32. Urbana: University of Illinois, 1926. Pp. 33-49.

The Experimental Coefficient.—It is frequently important to know whether or not measurements for a given group indicate that the given group is clearly superior to another group similarly measured and, if so, whether the superiority is such that it might always be anticipated or is such as might occur as the result of chance or random sampling. Sometimes it is important to know, when the obtained difference is not significant, how nearly the difference approaches to significance. The experimental coefficient is used for this purpose, reducing the measures of central tendency and of dispersion to a single measure which may be interpreted readily.

When the experimental coefficient has a value of 0.5, the chances are approximately 11 to 1 that the difference is statistically significant. They rise to approximately 369 to 1 when the value of the experimental coefficient reaches 1.0 and to about 67,000 to 1 when it reaches 1.5. An experimental coefficient of 1.0 is considered to indicate a significant difference.

An example of the use of the experimental coefficient is given in Table VII. These measures were computed from data included in Table I, page 62. The quadrants are numbered from qua-

drant one, the upper right-hand quadrant, in a counterclockwise direction.

TABLE VII.—EXPERIMENTAL COEFFICIENTS EVALUATING DIFFERENCES IN CITIZENSHIP DUE TO DIFFERENCES IN MENTAL ABILITY AND IN EXTROVERSION-INTROVERSION

Quadrant	2	3	4
1	0.81	0.79	0.30
2	0	0.47
3	0.44

It appears that differences in citizenship among pupils with differing mental abilities and differing extroversion-introversion characteristics are not such as may always be anticipated. The largest value of the experimental coefficients is only 81 per cent of what it should be to denote a statistically significant difference.

Readings for Further Study

1. H. E. GARRETT, *Statistics in Psychology and Education*. Pp. 210-212.
2. E. A. LINCOLN, The Insignificance of Significant Differences, *Journal of Experimental Education*, 2: 288-290 (1934).

The Error-analysis Chart.—So far, the discussion has been limited to a consideration of means of analyzing total scores yielded by measuring instruments. Mention should be made of a valuable method for securing information leading to improvement of classroom teaching based on a tabulation of the errors made by pupils responding to the items of a test. This device is known as an error-analysis chart. Table VIII presents such a chart, involving the errors made by 25 pupils on a 20-item test. The pupil identifications, usually pupils' names, are listed in a column at the left, and the numbers of the test items in a row across the top of the table. Each error made by each pupil is placed in the column headed by the item numbers of items on which errors were made. Totaling the errors for each pupil, across the table, and subtracting from the total possible score gives the individual pupil's achievement score. Totaling all the errors in any column yields information as to the types of prob-

lems missed. Analyzing the total errors on all items yields information leading to review and to remedial teaching.

TABLE VIII.—ERROR-ANALYSIS CHART ON AN ARITHMETIC TEST OF 20 ITEMS
N = 25

Pupil No.	Item Number																				Total errors per pupil
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1		x	x	x	x	x	x	x	x	x				x	x	x	x		x		14
2			x					x	x												3
3																					0
4		x	x			x				x	x	x	x				x			x	9
5			x						x									x			3
6	x	x	x							x			x		x	x	x		x		9
7							x						x	x			x		x		5
8							x			x			x				x				5
9												x						x			4
10							x		x									x		x	4
11										x						x				x	3
12			x						x												2
13						x	x			x		x						x	x	x	7
14	x		x				x	x	x	x			x	x	x			x		x	11
15													x	x	x	x	x				4
16													x						x		2
17		x										x	x				x	x		x	7
18										x			x					x			3
19		x		x				x								x				x	5
20								x	x	x	x			x		x	x	x		x	9
21		x							x		x			x	x						5
22			x					x	x		x				x	x				x	7
23										x	x			x							5
24																				x	1
25	x					x	x	x				x			x					x	7
Total errors per item	3	6	8	2	1	4	9	8	8	12	2	5	9	4	14	6	11	5	5	12	134

Readings for Further Study

1. M. E. BROOM, The Need for Statistics in Education, *High School Teacher*, 9: 300-304 (1933).

2. M. E. BROOM, *Statistics and the Classroom Teacher, High School Teacher*, 8: 13-15 (1932).
3. M. E. BROOM, *Educational Statistics for Beginning Students*. Pp. 171-186.
4. H. C. LEHMAN and P. A. WITTY, *Statistics Show, Journal of Educational Psychology*, 19: 175-184 (1928).
5. L. C. PRESSEY and S. L. PRESSEY, *Methods of Handling Test Scores*. Yonkers-on-Hudson, New York: World Book Company, 1926.

Exercises

On the following pages are given a number of practical problems, each based on data obtained in classroom situations. Only by thinking clearly about the meaning of statistical data can any real understanding develop concerning the significance of statistical materials. The student is urged to supplement these exercises by readings in professional journals reporting results of statistical investigations involving the use of educational measurements.

1. The Seashore Measures of Musical Talent appraise six basic capacities which underlie general music ability: pitch, intensity, time, rhythm, consonance, and tonal memory. The following table presents measures of relationship obtained by intercorrelating scores yielded by this test.¹

MEASURES OF RELATIONSHIP BETWEEN THE PAIRED SERIES OF SCORES YIELDED BY THE SEASHORE MEASURES OF MUSICAL TALENT

$N = 82$

(The probable errors vary from 0.062 to 0.074)

	Intensity	Time	Consonance	Memory	Rhythm
Pitch.....	0.39	0.29	0.28	0.41	0.34
Intensity.....	0.17	0.30	0.35	0.08
Time.....	0.31	0.24	0.32
Consonance.....	0.36	0.18
Memory.....	0.38

a. Why is it desirable that all of these coefficients should be positive in value?

b. Is it desirable that these coefficients be low in value? Why?

c. Explain the differences in the values of the different coefficients in the table above.

2. The constancy of the intelligence quotient is of interest to educators, since if it is not constant, the I.Q. cannot be used safely in determining mental age at times subsequent to the obtaining of the data. The following table presents information concerning I.Q. differences discovered by the

¹ M. E. BROOM, A Note concerning the Seashore Measures of Musical Talent, *School and Society*, 30: 275 (Aug. 24, 1929).

Otis Self-administering Test of Mental Ability, Higher Examination, over a five-month interval.¹

SUMMARY OF THE DIFFERENCES IN INTELLIGENCE QUOTIENTS ON TWO TESTINGS
 $N = 212$

I.Q. difference	N	Range		Md	Q
		From	To		
Same I.Q.	13				
Gains.....	143	1	22	5.3	3.7
Losses.....	56	1	16	4.6	2.6

Interpret the facts given in the table above, and attempt to explain their causes.

3. A silent reading test was administered to 4,435 pupils in six half grades in a public school system in October, 1936. The table on p. 76 presents summaries of the different score series by race and grade groups.²

a. Is there any definite trend, evidenced by the averages, indicating increasing achievement from the low fifth grade through the high seventh grade? Are the gains uniform for the different race groups from grade to grade? Are the achievement levels uniform within each grade level for the different sex and race groups?

b. Does either of the averages, mean and median, consistently exceed the other among the different groups of pupils? What is the significance of such differences as exist?

c. What facts are shown by the standard deviations? By the ranges?

d. How could the findings presented in this table be shown graphically? In your opinion, would a graph present the findings more clearly than a table? Why?

4. Achievement differences among groups can be interpreted in terms of critical ratios, in part, at least. The table on p. 77 presents data concerning differences between race groups. It is based on the summary measures included in the table given in Exercise 3.³

Consider each paired race group critical ratio. Is there any sex and race group that is consistently different from any other different race group of the same sex? Which? Can significant conclusions be based on the data in this table? If so, what are they?

¹ M. E. BROOM, How Constant Is the I.Q. Yielded by the Otis Self-administering Test of Mental Ability? *Journal of Educational Research*, 22: 54 (June, 1930).

² C. W. WEBB, Surveying Reading Results and Locating Opportunities for Modification, *El Paso Schools Standard*, 15: 31 (January, 1938).

³ WEBB, *op. cit.* P. 33.

76 MEASUREMENTS IN THE ELEMENTARY SCHOOL

SUMMARY OF SCORES, ALL GRADES, OCTOBER, 1936, TESTING

Grade	Race	Sex	N	Range		Md	M	σ
				From	To			
H 7	Anglo-American	Boys	125	49	98	82.9	81.0	12.5
H 7	Anglo-American	Girls	101	42	97	82.8	80.1	12.2
H 7	Spanish-American	Boys	151	34	93	69.3	68.1	14.3
H 7	Spanish-American	Girls	167	38	97	70.5	69.1	13.1
L 7	Anglo-American	Boys	180	49	96	82.0	80.5	10.7
L 7	Anglo-American	Girls	148	40	97	83.4	82.0	10.1
L 7	Spanish-American	Boys	177	28	96	63.4	59.2	11.7
L 7	Spanish-American	Girls	181	30	99	68.6	60.3	11.1
L 7	Negro	Boys	13	25	91	76.3	79.1	9.2
L 7	Negro	Girls	11	78	97	85.7	85.4	8.9
H 6	Anglo-American	Boys	118	18	94	70.7	67.3	13.4
H 6	Anglo-American	Girls	112	40	95	76.0	73.0	11.7
H 6	Spanish-American	Boys	223	16	96	55.8	56.5	12.2
H 6	Spanish-American	Girls	179	19	89	56.7	57.6	11.4
L 6	Anglo-American	Boys	209	21	94	64.2	63.4	12.7
L 6	Anglo-American	Girls	194	35	92	68.4	67.0	11.6
L 6	Spanish-American	Boys	229	11	94	50.3	51.9	12.0
L 6	Spanish-American	Girls	247	16	88	51.8	52.4	12.0
L 6	Negro	Boys	17	56	96	73.0	71.6	11.4
L 6	Negro	Girls	13	5	87	75.7	70.7	14.3
H 5	Anglo-American	Boys	112	20	91	61.5	60.9	12.6
H 5	Anglo-American	Girls	101	28	96	63.1	63.1	12.0
H 5	Spanish-American	Boys	275	10	91	43.7	44.9	12.3
H 5	Spanish-American	Girls	211	15	83	47.0	47.8	11.4
L 5	Anglo-American	Boys	140	8	80	47.7	44.9	12.8
L 5	Anglo-American	Girls	116	20	82	53.9	55.2	12.6
L 5	Spanish-American	Boys	319	3	78	34.5	36.1	11.9
L 5	Spanish-American	Girls	330	5	70	37.8	37.8	11.4
L 5	Negro	Boys	17	18	65	48.5	44.5	12.0
L 5	Negro	Girls	19	22	60	42.0	43.0	10.0

CRITICAL RATIOS FOR ACHIEVEMENT IN READING FOR CERTAIN SEX AND RACE GROUPS

Grade	Groups compared	<i>EC</i>	Groups favored
H 7	Anglo-American boys vs. Spanish-American boys	2.88	Anglo-Americans
H 7	Anglo-American girls vs. Spanish-American girls	2.45	Anglo-Americans
L 7	Anglo-American boys vs. Negro boys	0.19	Anglo-Americans
L 7	Anglo-American girls vs. Spanish-American girls	6.57	Anglo-Americans
L 7	Negro girls vs. Anglo-American girls	0.43	Negroes
L 7	Negro girls vs. Spanish-American girls	1.93	Negroes
L 7	Negro boys vs. Spanish-American boys	2.65	Negroes
L 7	Anglo-American boys vs. Spanish-American boys	6.65	Anglo-Americans
H 6	Anglo-American girls vs. Spanish-American girls	3.96	Anglo-Americans
H 6	Anglo-American boys vs. Spanish-American boys	2.62	Anglo-Americans
L 6	Anglo-American boys vs. Spanish-American boys	3.39	Anglo-Americans
L 6	Negro boys vs. Spanish-American boys	2.46	Negroes
L 6	Negro boys vs. Anglo-American boys	1.00	Negroes
L 6	Anglo-American girls vs. Spanish-American girls	4.64	Anglo-Americans
L 6	Negro girls vs. Anglo-American girls	0.24	Negroes
L 6	Negro girls vs. Spanish-American girls	1.65	Negroes
H 5	Anglo-American girls vs. Spanish-American girls	3.86	Anglo-Americans
H 5	Anglo-American boys vs. Spanish-American boys	6.94	Anglo-Americans
L 5	Anglo-American girls vs. Negro girls	1.74	Anglo-Americans
L 5	Negro girls vs. Spanish-American girls	0.84	Negroes
L 5	Anglo-American girls vs. Spanish-American girls	7.34	Anglo-Americans
L 5	Anglo-American boys vs. Negro boys	0.46	Anglo-Americans
L 5	Negro boys vs. Spanish-American boys	1.01	Negroes
L 5	Anglo-American boys vs. Spanish-American boys	2.49	Anglo-Americans

6. Four individual tests were administered to pupils in an elementary school. Not all of the pupils responded to all of the tests. The following table presents the summary data concerning mental age.¹

SUMMARY OF THE DISTRIBUTIONS OF MEASURES OF MENTAL AGE

Test	N	Range		M	σ
		From	To		
Kohs.....	89	5-6	19-8	12-11	2-10
Porteus.....	82	5-0	18-0	12-10	2-6
Herring-Binet.....	89	5-10	19-8	12-10	3-5
Stanford-Binet.....	37	5-3	16-8	12-6	2-7

Interpret the data given in the table above.

6. The Kohs Block Design Test, the Herring Revision and the Stanford Revision (1916) of the Binet-Simon Tests, and the Porteus Maze Test are individual tests purporting to yield a measure of mental ability. The table below presents intercorrelations indicating the relationship between paired measures yielded by these tests.²

MEASURES OF RELATIONSHIP BETWEEN THE PAIRED SERIES OF MENTAL AGES YIELDED BY FOUR INDIVIDUAL TESTS OF MENTAL ABILITY
(N varies from 14 to 82; the probable errors, from 0.02 to 0.08)

Test	Porteus	Herring-Binet	Stanford-Binet
Kohs.....	0.789	0.868	0.659
Porteus.....	0.727	0.614
Herring-Binet.....	0.748

a. What important characteristic of a test is evidenced by the data in the table above? Is the level indicated in the table satisfactory for all four tests?

b. Comment on your findings as to differing coefficient values in the table.

7. Three standardized tests and two teacher-made, objective tests of silent reading comprehension were administered to pupils in grade 6 in an elementary school during April, 1937. Intercorrelations based upon the resulting scores are given in the table on page 79.³

¹ M. E. BROOM, *The Validity of Four Individual Tests of Mental Ability*, *Los Angeles Educational Research Bulletin*, 8 (No. 7): 10 (June, 1929).

² M. E. BROOM, *ibid.*, p. 10.

³ M. E. BROOM, A. MCGURN, and M. TRACHSEL, *A Silent Reading Test for Upper Elementary Grades*. *El Paso Schools Standard*, 15 (No. 1): 9 (October, 1937).

Test	Nelson	Sangren-Woody	El Paso I	El Paso II
Los Angeles Elementary.....	0.857	0.697	0.806	0.783
Nelson.....	0.836	0.847	0.880
Sangren-Woody.....	0.818	0.853
El Paso I.....	0.913

a. How many validity coefficients appear in the table above? How many reliability coefficients?

b. Name the tests concerning which validity evidence is given.

c. Do the reliability and validity coefficients in the table above indicate that the tests are satisfactory in these characteristics? Why?

d. By listing all of the validity coefficients for each test and then finding the median coefficient value, determine the order of excellence of the tests with reference to validity.

8. The following table presents data on a unit on *Texas Today*, taught to pupils in grade 6 in a public school during the first semester of 1937-1938. The tests used were teacher-made objective examinations.¹

SUMMARY OF ACHIEVEMENT SCORES ON UNIT XII

N = 29

Test	Range		Md	M	σ	EC	r
	From	To					
I (as pretest).....	3	17	12.42	11.92	4.41	} 7.04	} 0.71
I (as $\frac{1}{2}$ post test)....	24	40	35.28	34.25	4.3		
II (as $\frac{1}{2}$ post test)...	27	40	33.8	33.85	3.85		

a. Test I was used as the pretest and as one-half the post test, to obtain evidence as to gains resulting from instruction. What evidence is contained in the table above that there were achievement gains during the period of instruction? (Consider range, the averages, the standard deviation, and the experimental coefficient.)

b. Test II is given to validate the gains resulting from instruction, indicating that such gains were actual gains in information, not resulting from memory of the test. What evidence is contained in the table above that the two tests were reasonably equivalent? What evidence that gains were valid ones, not resulting from memory of Test I?

¹ DOROTHY EMIG, A Study of Unit Instruction in Sixth Grade Social Science, p. 61, unpublished Master's thesis, State Teachers College, Alpine, Tex., 1938.

c. What data from the table on page 79 might be presented effectively in graphical form?

9. The following table presents data for 12 units, including the measures that indicate significant gains due to instruction. These units were taught during the school years 1935-1936, 1936-1937, and 1937-1938 to pupils in grade 6 in a public school.¹

MEASURES INDICATING SIGNIFICANCE OF GAINS DUE TO INSTRUCTION

Measure	N	Range		Md value	No. indicating significant gains in achievement
		From	To		
EC	12	4.79	9.56	7.54	12
r	12	0.55	0.90	0.71	12

Interpret these findings, keeping in mind the interpretation of Exercise 8, and explain why they show that the unit method of instruction is productive of pupil achievement.

10. The table below presents data summarizing intelligence quotients for boys and girls separately for three mental tests. These tests were given to high eighth-grade pupils in a public-school system during the school years 1935-1936, 1936-1937, and 1937-1938.

SUMMARIES OF INTELLIGENCE QUOTIENTS YIELDED BY THREE MENTAL TESTS BY SEX GROUPS

Test	Sex group	N	Range		Md	M	σ	EC
			From	To				
Miller.....	Boys	297	61	176	110.4	111.3	23.8	0.28
	Girls	333	49	180	106.3	109.7	27.8	
McCall.....	Boys	837	59	162	99.2	101.3	20.8	0.87
	Girls	869	54	182	101.4	103.7	21.6	
Henmon-Nelson.....	Boys	639	64	158	101.2	102.5	16.3	0.08
	Girls	654	61	145	101.1	102.7	16.9	

a. Compare the mean and the median for each group of pupils. Does either consistently exceed the other in value? Which is the better measure of central tendency?

b. Which of the three tests exhibits the greater homogeneity?

c. What is the significance, if any, of differences in the ranges for each test? How many σ distances are included in each range?

¹ *Ibid.*, pp. 63, 65.

d. Do any of the experimental coefficients indicate a significant sex difference? Is either sex group consistently favored by the obtained differences? What is the significance of this for mental testing?

11. The following table presents coefficients of correlation indicating the relationship existing between scores yielded by six standardized civics tests. These tests were administered to 108 pupils in a public school during 1932-1933.¹

MEASURES OF RELATIONSHIP BETWEEN PAIRED SCORES YIELDED BY SIX
STANDARDIZED CIVICS TESTS
N = 108

Test	Hill CI	Brown- Woody A	Brown- Woody B	Bur- ton A	Bur- ton B	Almack I	Almack II	Ma- gruder A	Ma- gruder B
Hill CA.....	0.480	0.485	0.517	0.488	0.496	0.526	0.419	0.470	0.491
CI.....		0.639	0.529	0.513	0.524	0.597	0.604	0.498	0.468
Brown-Woody A....			0.756	0.574	0.562	0.523	0.647	0.738	0.695
B.....				0.577	0.580	0.590	0.637	0.695	0.662
Burton A.....					0.522	0.493	0.492	0.468	0.474
B.....						0.559	0.446	0.560	0.472
Almack I.....							0.632	0.713	0.680
II.....								0.734	0.736
Magruder A.....									0.812

NOTE.—The probable errors vary from 0.02 to 0.05. The letters and Roman numerals after the names of certain tests indicate the forms used. The Hill Civic Information and Civic Attitudes tests have only one form and were used once only.

a. How many validity coefficients appear in the table above? How many reliability coefficients?

b. Each of these tests, by virtue of the procedures taken during its standardization, has a definite claim to validity and reliability. Are these equally valid? Which tests tend to validate each other? Which tests are invalid, judging from the data in the table? What can you say concerning the reliability of these tests?

12. A battery of achievement tests was given to 100 pupils in a public school in October, 1936. One year later, in October, 1937, a duplicate form of the same achievement test was given to the same 100 pupils. During March, 1936, these pupils were given the Henman-Nelson Tests of Mental Ability and a personality test yielding measures of leadership traits, the Morris Trait Index L. The first table on page 82 summarizes the achievement

¹L. SMITH, A Study of Six Standardized Civics Tests Devised for Use in Secondary Schools, p. 37, unpublished Master's thesis, University of Southern California, 1933.

gains in point scores on the battery of achievement tests, total scores, for pupils grouped according to their placement above the medians for the personality and mental tests.¹ The second table presents experimental coefficients indicating the significance of differences in increased achievement over the period of one year for pupils grouped into the four quadrants on the basis of mental ability and personality differences, as evidenced by the tests.

SUMMARIES OF ACHIEVEMENT GAINS IN POINT SCORES YIELDED BY A
STANDARDIZED BATTERY OF ACHIEVEMENT TESTS DURING A PERIOD OF
ONE YEAR FOR 100 PUPILS GROUPED ACCORDING TO DIFFERENCES
IN PERSONALITY AND IN MENTAL ABILITY

Trait Index L

Below the median *Md* Above the median

Mental ability	Above the median	Quadrant 2 $N = 14$ Range = -2 to 57 $M = 21.64$ $\sigma = 18.68$	Quadrant 1 $N = 36$ Range = 1 to 79 $M = 31.44$ $\sigma = 16.80$
	<i>Md</i>	Quadrant 3 $N = 36$ Range = -15 to 64 $M = 14.22$ $\sigma = 16.91$	Quadrant 4 $N = 14$ Range = 7 to 70 $M = 21.64$ $\sigma = 16.66$
	Below the median		

EXPERIMENTAL COEFFICIENTS INDICATING THE SIGNIFICANCE OF DIFFERENCES IN ACHIEVEMENT GAINS BY PUPILS GROUPED ACCORDING TO DIFFERENCES IN PERSONALITY AND IN MENTAL ABILITY

Quadrants	2	3	4
1	0.62	0.68	0.67
2	0.47	0
3	0.51

Interpret the meaning of the measures given in the two preceding tables. Summarize your findings with a conclusion as to the relative significance of mental ability and leadership traits in producing achievement.

EXAMINATION ON STATISTICAL PROCEDURES

Directions: Some of the statements below are true; some are false; and some are partially true and partially false. Place a check on the line at the left of the letter T before each true statement; a check on the line at the left of the letter F before each false statement; and a check on the line at the

¹ These data were furnished by Josephine M. Skiff, El Paso, Tex., Public Schools.

left of the letter N before each statement that is neither entirely true nor entirely false. If you do not know which response to check, skip the item entirely, and continue at once with the next one. **DO NOT GUESS!** Work as rapidly and as accurately as you can.

- | | | | | | | |
|--------------------------|---|--------------------------|---|--------------------------|---|--|
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 1. A class interval means a portion of the range in which no cases fall. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 2. No matter what assumed mean is chosen, the same resultant will be obtained if the computation is accurate. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 3. The mean of a frequency distribution is always a point on the linear scale of values included in its range. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 4. The median takes account of the number but not of the relative value of measures in a frequency distribution. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 5. The mean and the median have the same numerical value when the distribution is normal. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 6. The mean is based not only on all the measures in a frequency distribution but also on their exact magnitude. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 7. A distribution has been sufficiently described when its average, or central tendency, and its range have been stated. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 8. One standard deviation marked off on either side of the mean in any distribution includes exactly the middle two-thirds of the given cases. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 9. Each successive standard deviation distance from the mean includes an equal number of cases. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 10. All percentile curves, ogives, tend to have roughly the same characteristic shape or curve. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 11. The median should be used to indicate central tendency in a distribution when it is desirable to avoid the influence of extreme or erroneous values. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 12. N , in statistical formulas, is always the total frequency. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 13. The range of different values included in a given class is the class interval. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 14. The general arrangement on a chart or graph should proceed from left to right and ordinarily from bottom to top. |
| <input type="checkbox"/> | T | <input type="checkbox"/> | F | <input type="checkbox"/> | N | 15. The algebraic sum of the deviations from the median is always zero. |

- _____ T _____ F _____ N 16. The magnitude of the range depends upon the values of all the measures in a frequency distribution.
- _____ T _____ F _____ N 17. A frequency table hides somewhat in its mass of figures the relations existing between the various data.
- _____ T _____ F _____ N 18. In the long method of computing the mean and various other statistical expressions, use is made of an assumed or guessed mean.
- _____ T _____ F _____ N 19. A histogram or column diagram is a common type of frequency curve.
- _____ T _____ F _____ N 20. A frequency table consists of one column of numbers that indicates the limits of the various classes into which the individual measures have been grouped and a second that shows the number of cases in each class.
- _____ T _____ F _____ N 21. The frequency polygon is constructed by placing at the upper limit of each class and connecting with straight lines a series of points, each one of which is at the mid-point of the class and at a height equal to the frequency in the class.
- _____ T _____ F _____ N 22. In statistical studies involving the determination of relationship, two or more variables or series of measures are concerned.
- _____ T _____ F _____ N 23. A numerical index of correlation ordinarily has only one meaning: It shows the actual degree of relationship existing between the cases upon which it is based.
- _____ T _____ F _____ N 24. Positive correlation means that as one of the two paired variables increases, the other tends to decrease proportionately.
- _____ T _____ F _____ N 25. The calculation of the product-moment coefficient of correlation requires the same computation steps used in calculating the mean and the standard deviation, plus some additional work.
- _____ T _____ F _____ N 26. It is quite easy to interpret the meaning of a coefficient of correlation in ordinary thought and language.
- _____ T _____ F _____ N 27. The coefficient of correlation is a minimum expression of the amount of

relationship that exists between two paired series of numerical facts.

- _____ T _____ F _____ N 28. The fact that two series of measures are correlated does not indicate that either causes or is in any way responsible for the other.
- _____ T _____ F _____ N 29. The coefficient of reliability is merely the coefficient of correlation between two applications to the same individuals of the same measuring instrument
- _____ T _____ F _____ N 30. The experimental coefficient is a critical ratio.
- _____ T _____ F _____ N 31. An experimental coefficient of 1.5 indicates that the difference is such as might occur as the result of chance or random sampling.
- _____ T _____ F _____ N 32. The minimum value for validity coefficients is positive 0.50.
- _____ T _____ F _____ N 33. Reliability coefficients should have a minimum value of positive 0.90, if the tests that such coefficients describe are to be considered sufficiently reliable.
- _____ T _____ F _____ N 34. The experimental coefficient is applicable only in case the significance of a difference (as $M_1 - M_2$) is in question.
- _____ T _____ F _____ N 35. The coefficient of correlation may yield misleading information, if the given data are not well summarized by a linear trend line.
- _____ T _____ F _____ N 36. If there is no connection whatever between two paired series of variables, r is 0 in value.
- _____ T _____ F _____ N 37. The coefficient of correlation may be relied upon to give a complete analysis of relationships in a given problem situation.
- _____ T _____ F _____ N 38. Tertiles, quartiles, deciles, and other percentile points may be computed by the same general formula, with appropriate minor changes.
- _____ T _____ F _____ N 39. The significance of a coefficient of correlation may be invalidated by accidental relationships or by defects through a paucity of data.
- _____ T _____ F _____ N 40. The measures in all classes in a frequency table are assumed in each class to be uniformly distributed throughout it.

- T F N 41. The ogive is a cumulative frequency curve.
- T F N 42. The quartile deviation is weak as a measure of dispersion because it is not a distance from any average.
- T F N 43. The most commonly used measure of variability is the range.
- T F N 44. A frequency polygon touches the base line at class limits.
- T F N 45. The point on the scale of values about which the measures composing a frequency distribution tend to group themselves is called the central tendency.
- T F N 46. Correlation is always defined as the correspondence between two paired series of facts.
- T F N 47. The product moment coefficient of correlation deals with straight-line relationship only.
- T F N 48. A frequency curve composed of a series of adjacent rectangles is called a summary graph.
- T F N 49. The experimental coefficient is used in connection with errors of measurement.
- T F N 50. The chief purpose in using a prepared correlation chart is to insure neatness in assembling and presenting the work.

CHAPTER V

STANDARDIZED VERSUS TEACHER-MADE OBJECTIVE EXAMINATIONS

The standardized test is regarded by many educators as being a more scientific instrument than any classroom teacher can construct. Some teachers seem to feel that it possesses unusual virtues as a measuring instrument merely because it is a standardized instrument. Such beliefs are entirely mistaken.

Almost a decade ago, Ruch wrote as follows concerning this matter:

The best of existing standard tests do represent a degree of refinement not possible without extended experimentation. But the rank-and-file of such tests are readily equaled or bettered by the teacher who has mastered a little theory of measurement and who is seriously intent upon building valid and reliable examinations.¹

The purpose of this chapter is to present evidence concerning this matter.

The test maker must have certain knowledge concerning the construction of examinations and a command of the skills needed in test building. It is entirely possible for a test maker to develop an objective examination, in which a variety of types of problems may be included, arranged in the judgment of the test constructor in the order of their probable difficulty. Properly constructed, such an objective test may be scored readily and in a minimum of time. Certainly the cost of mimeographing such a nonstandardized, objective test, similar in length to standardized tests devised to measure the same thing or things that are measured by the objective test, will be a relatively small fraction of the cost of the standardized tests. Of the essential characteristics of a good examination, the test constructor can be reasonably assured, then, of all being considered, except validity, reliability, dis-

¹ G. M. RUCH, *The Objective or New-type Examination*, Introduction, p. iv.

crimination, and freedom from coaching. The best method for securing a valid examination is the scrutiny of all items proposed for inclusion in the test with a view to determining whether or not each is within the proposed scope of the test and also whether or not it is subject matter commonly stressed in the best teaching in the given subject and school grade for which the test is being prepared. This may imply the examination of superior textbooks and curriculums in the given subject and for the given grade; very definitely it requires that the constructor know the subject matter and good teaching methods and procedures in the given subject and school grade. To insure reliability in the test requires (1) a reasonably long measuring instrument, (2) with items arranged approximately in order of difficulty, with some exercises presented early in the test so easy that every pupil responding to it can do some items successfully and others so difficult that no pupil can do all of the items successfully in the time allowed. Of these two things, length is the more important. If the items in the test are valid, and these precautions have been followed, the test will have sufficient reliability. Furthermore, these precautions in the construction will go far in insuring a test that will be discriminative and also reasonably free from coaching effects.

Of the several characteristics, validity, reliability, and objectivity are the most important. The last is cared for in the construction of the test. Evidence from a number of studies is available, showing the comparative validity and reliability of several standardized tests and of one or more teacher-made objective examinations. Brief presentations of this evidence, with explanatory comments, follow.

Two Studies of Silent Reading Tests.—Both of these studies were concerned with the measurement of silent reading comprehension in the middle elementary grades. In each instance, several standardized tests were used, together with one or more teacher-made, objective tests of silent reading comprehension. The latter instruments were devised to measure several types of reading skills, presented in a variety of ways in the tests. The instruments prepared by the experimenters, using techniques commonly employed during the standardization of tests, were similar in content and format to, but probably cruder than, certain of the standardized reading tests.

In the first study, the following tests were used:

- The Stanford Reading Test, Form A
- The Sangren-Woody Reading Test, Form A
- The Los Angeles Elementary Reading Test, Form 1
- The Public School Achievement Tests, Reading Test, Form
- The Haggerty Reading Examination, Sigma 3, Form A
- An unstandardized, teacher-made objective reading test

Each of these tests was administered once to pupils in grade 6 in an elementary school. Because of absence, not all of the pupils responded to all of the tests. The data are presented in Table IX.

TABLE IX.—VALIDITY COEFFICIENTS FOR SIX SILENT READING TESTS¹

Tests	Sangren-Woody	Los Angeles	Orleans	Haggerty Σ3	Unstandardized ²	Average of 5 standard tests ³
Stanford.....	0.66	0.70	0.74	0.86	0.76	0.85
Sangren-Woody.....	0.69	0.71	0.74	0.82	0.85
Los Angeles Elementary.....	0.66	0.76	0.71	0.82
Orleans (Public School Achievement Test).....	0.75	0.69	0.74
Haggerty Σ3.....	0.80	0.89
Unstandardized.....	0.81

NOTE.—The probable errors of the coefficients vary from 0.02 to 0.05; the numbers of cases, from 66 to 70. The data were obtained in grade 6 of the Chula Vista, Calif., Elementary School during the school year 1928-1929.

¹ M. E. BROOM, JOSEPHINE DOUGLAS, and MARION RUDD, On the Validity of Silent Reading Tests, *Journal of Applied Psychology*, 15: 35-38 (February, 1931).

² This test was constructed by students in a class in educational measurements taught by the writer in the San Diego State College during the second semester of the school year 1928-1929.

³ The standard score unit was $\frac{1}{2}\sigma$. Zero was taken at -3σ ; -2.9σ was given a value of 1; -2.8σ , a value of 2; etc. The standard score values of scores earned by any one pupil on all five standardized tests were averaged. In three or four instances, as few as three such derived scores yielded by standardized tests were used. Absence from the school limited the number of tests taken by these pupils.

If the criterion of the median value validity coefficient is applied in determining the relative merit of the tests, the Haggerty Reading Examination, Sigma 3, and the teacher-made objective test of reading comprehension are equally valid and in first rank in excellence with a median validity coefficient of positive 0.76. If the criterion of correlation with the average of the standardized test scores is applied, all but one of the standardized tests exceed

the teacher-made objective test of reading comprehension in excellence as to validity. It should be kept in mind, however, that each of the standardized tests contributed to the average score used as the criterion in this instance, whereas the unstandardized test was not included in obtaining the average score. There necessarily was some self-agreement between the average and individual standardized test score series, increasing the value of the coefficient in each instance in which a standardized test was involved. Taking this fact into account, it becomes obvious that the teacher-made, objective test has validity as good as that of the average of the standardized tests of reading comprehension.

In the second study, the following tests were used:

The New Stanford Reading Test, Forms V and W

The Sangren-Woody Reading Test, Forms A and B

The Ingraham-Clark Diagnostic Reading Tests, Intermediate Test, Part II, Forms 1 and 2

The Los Angeles Elementary Reading Test, Forms 1 and 2

Two unstandardized, teacher-made objective tests of reading comprehension, each of which was administered twice

The individual forms of the standardized tests were given once only. Because either two forms of the standardized tests were given, or the unstandardized tests were given twice to the pupils, it was possible to determine both the reliability and the validity of the tests. The tests were given to pupils in grades 4 and 5 in an elementary school; and because of absence from school, not all of the pupils responded to all of the tests. The data are given in Table X.

If, again, there is applied the criterion of the median coefficient value in determining validity, the order of excellence is (1) The New Stanford Reading Test (0.715); (2) Unstandardized Test I (0.700); (3) Unstandardized Test II and The Los Angeles Elementary Reading Test (both with 0.685); (5) The Sangren-Woody Reading Test (0.680); and (6) The Ingraham Clark Diagnostic Reading Test (0.650). These six different instruments possess validity in excess of the minimum recommended value, positive 0.60. The order of excellence as determined by the reliability coefficients is (1) The New Stanford Reading Test, (2) Unstandardized Test II, (3) The Los Angeles Elementary Reading Test, (4) Unstandardized Test I and The Sangren-Woody Reading

Test, and (6) The Ingraham-Clark Diagnostic Reading Test. It should be noted that all of the reliability coefficients are below the desired minimum value, positive 0.90.

The evidence from both of these investigations indicates that it is possible for teachers to construct objective tests of reading

TABLE X.—MEASURES OF THE VALIDITY AND OF THE RELIABILITY OF CERTAIN SILENT READING TESTS¹

Tests	New Stanford W	Sangren-Woody A	Sangren-Woody B	Ingraham-Clark I	Ingraham-Clark II	Los Angeles Elementary 1	Los Angeles Elementary 2	Unstandardized I(1)	Unstandardized I(2)	Unstandardized II(1)	Unstandardized II(2)
New Stanford V	0.85	0.73	0.70	0.76	0.65	0.78	0.74	0.64	0.75	0.72	0.68
W	0.72	0.68	0.71	0.70	0.78	0.74	0.64	0.72	0.70	0.66	
Sangren-Woody A		0.79	0.60	0.63	0.68	0.61	0.69	0.78	0.73	0.68	
B			0.58	0.54	0.65	0.57	0.58	0.70	0.71	0.60	
Ingraham-Clark I				0.68	0.56	0.75	0.60	0.73	0.69	0.68	
II					0.65	0.67	0.55	0.72	0.63	0.61	
Los Angeles Elementary 1						0.80	0.63	0.74	0.75	0.68	
2							0.62	0.71	0.74	0.69	
Unstandardized I(1)								0.79	0.64	0.60	
(2)									0.82	0.78	
Unstandardized II(1)										0.84	

NOTE.—The probable errors of the coefficients vary from 0.017 to 0.044; the numbers of cases, from 118 to 120. The data were obtained in grades 4 and 5 of the National City, Calif., elementary schools during the school year 1930-1931.

¹ B. H. GIBSON, A Study of the Use of Teacher-made Tests in the Improvement of Pupil Achievement, unpublished Master's thesis, pp. 69f., University of Southern California, 1931.

comprehension that will be as valid and as reliable as the better standardized silent reading tests which are available for classroom use.

A Study of Algebra Tests.—This investigation was carried out in a junior high school, with the tests being given to pupils in grade 9. The tests used included the following measuring instruments:

The Douglass Standard Survey Test for Elementary Algebra, Test I, Forms A and B

The Illinois Standardized Algebra Test, administered twice

The Columbia Research Bureau Algebra Test, Test I, Forms A and B

Two unstandardized, teacher-made algebra tests, each of which was administered twice

The three standardized tests have been described and recommended by a number of authorities¹ for use in an elementary course in algebra. The first of the teacher-made examinations

TABLE XI.—MEASURES OF THE VALIDITY AND RELIABILITY OF CERTAIN ALGEBRA TESTS¹

Tests	Douglass B	Illinois (1)	Illinois (2)	Columbia I-A	Columbia I-B	Unstandardized I(1)	Unstandardized I(2)	Unstandardized II(1)	Unstandardized II(2)
Douglass A.....	0.885	0.680	0.659	0.773	0.786	0.434	0.732	0.705	0.776
B.....		0.619	0.593	0.818	0.784	0.494	0.596	0.710	0.741
Illinois (1).....			0.807	0.652	0.627	0.453	0.621	0.690	0.565
(2).....				0.666	0.544	0.463	0.493	0.573	0.576
Columbia I-A.....					0.802	0.583	0.654	0.672	0.717
I-B.....						0.421	0.605	0.692	0.690
Unstandardized I(1).....							0.676	0.301	0.381
I(2).....								0.497	0.572
Unstandardized II(1).....									0.855

NOTE.—The probable errors of the coefficients vary from 0.014 to 0.060; the number of cases is 110; 49 pupils in grade 9-B, first-semester algebra; 61 pupils in grade 9-A, second-semester algebra. The data were obtained from the Memorial Junior High School, San Diego, Calif., second semester 1929-1930.

¹ B. H. BROOM, *A Study of Ninth Grade Algebra Tests*, p. 21, unpublished Master's thesis, University of Southern California, 1932.

¹ C. W. ODELL, *Educational Measurement in High School*, pp. 214-219.

G. M. RUCH and G. D. STODDARD, *Tests and Measurements in High School Instruction*, pp. 77-80, Yonkers-on-Hudson, New York: World Book Company, 1927.

E. W. TIEGS, *Tests and Measurements for Teachers*, pp. 383-384, Boston: Houghton, Mifflin Company, 1931.

G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.), pp. 416-418, New York: The Macmillan Company, 1928.

in algebra included a total of 75 items, presented in objective form and including exercises in solving equations, exercises in reduction of given expressions to a more convenient form, and algebraic formulas and algebraic terms. The second teacher-made examination was on problem solving, and it included 30 exercises. Six textbooks and drill books in elementary algebra were consulted by the investigator before the teacher-made examinations were prepared.¹

The data concerning the reliability and the validity of the five algebra tests are presented in Table XI.

The order of excellence as to validity, using the median validity coefficient value as the measure of validity, is as follows: (1) The Douglass Standard Survey Test for Elementary Algebra (0.708); (2) Unstandardized Algebra Test II (0.681); (3) The Columbia Research Bureau Algebra Test (0.669); (4) The Illinois Standardized Algebra Test (0.607); and (5) Unstandardized Algebra Test I (0.496). The order of excellence as to reliability, determined by the values of the reliability coefficients, is (1) The Douglass Standard Survey Test for Elementary Algebra, (2) Unstandardized Algebra Test II, (3) The Illinois Standardized Algebra Test, (4) The Columbia Research Bureau Algebra Test, and (5) Unstandardized Algebra Test I. Four of the five tests, excepting Unstandardized Test I, meet the minimum requirements as to validity on the average, 0.60. None of the five tests meets the minimum requirement as to reliability, 0.90, although one, the Douglass test, achieves 0.885. It is clear that in algebra, as in reading, teacher-made objective tests can be constructed to possess as good reliability and validity as do the better standardized tests. The evidence also shows that this result cannot be accomplished with every such objective test prepared by the classroom teacher.

A Study of Foods Tests.—The tests used in this investigation include

An Information Test on Foods, prepared by a test committee of the Illinois Home Economics Association

The King-Clark Food Test, devised by Florence King and Harold Clark

The Streater-Trilling Food Preparation Test, devised by Nina Streater and Mabel Trilling

¹ B. H. Broom, *A Study of Ninth Grade Algebra Tests*, pp. 8, 9.

A Ninth Grade Foods Test (unstandardized), prepared by Luella Kramer, Lawrence Bohnhoff, and Irma Shotola¹

Three teacher-made objective tests, designated as Tests A, B, and C.

These tests were given to girls in foods classes in grades 10 to 12 inclusive in two senior high schools. Each test was given twice, excepting Tests A and B, which were devised to be equivalent in total difficulty on the basis of the judgment of the test maker. These two tests included 19 recall items, 40 true-false

TABLE XII.—MEASURES OF THE VALIDITY AND OF THE RELIABILITY OF CERTAIN FOOD TESTS¹

Tests	King-Clark (2)	Illinois (1)	Illinois (2)	Streator-Trilling (1)	Streator-Trilling (2)	Kramer <i>et al.</i> (1)	Kramer <i>et al.</i> (2)	Unstandardized A	Unstandardized B	Unstandardized C (1)	Unstandardized C (2)
King-Clark (1).....	0.66	0.50	0.51	0.35	0.40	0.50	0.56	0.45	0.41	0.47	0.51
(2).....		0.55	0.58	0.36	0.41	0.55	0.60	0.50	0.54	0.57	0.58
Illinois (1).....			0.78	0.44	0.34	0.68	0.62	0.58	0.57	0.58	0.59
(2).....				0.56	0.44	0.64	0.71	0.60	0.56	0.64	0.82
Streator-Trilling (1).....					0.67	0.36	0.47	0.41	0.45	0.47	0.49
(2).....						0.40	0.48	0.44	0.45	0.45	0.50
Kramer, <i>et al.</i> (1).....							0.79	0.58	0.57	0.58	0.60
(2).....								0.60	0.61	0.69	0.71
Unstandardized A.....									0.86	0.61	0.63
B.....										0.67	0.71
C(1).....											0.90

NOTE.—The probable errors of the coefficients vary from 0.012 to 0.056; the number of cases is 116. The data were obtained from San Diego, Calif., secondary schools in April, 1931.

¹FAY V. PERRY, A Study of Foods Tests for Use in the Secondary School, p. 30, unpublished Master's thesis, University of Southern California, 1931.

items, 25 simple recall items, 5 rearrangement items, and 12 matching items in each examination. Test C included 100 items in four-answer, multiple-response form. The time for each of these three tests was 25 minutes.

The data are presented in Table XII.

¹From G. M. RUCH and G. D. RICE, *Specimen Objective Examinations*, pp. 283-287. Chicago: Scott, Foresman & Company, 1930.

The teacher-made objective tests are superior to the standardized tests in this study, both in reliability and in validity. The order of excellence as to validity, using the median coefficient value as the indicator of validity, is (1) The Kramer-Bohnhoff-Shotola Ninth Grade Foods Test (0.590); (2) Unstandardized Test C (0.585); (3) The Illinois Information Test on Foods and Unstandardized Test A (both with 0.580); (5) Unstandardized Test B (0.570); (6) The King-Clark Food Test (0.505); and (7) The Streator-Trilling Food Preparation Test (0.445). The order of excellence, in terms of the values of the reliability coefficients, is (1) Unstandardized Test C; (2) Unstandardized Tests A-B, provided it is assumed that these identical in form, but not necessarily equivalent, instruments may be considered duplicate forms; (3) The Kramer-Bohnhoff-Shotola Ninth Grade Foods Test; (4) The Illinois Information Test on Foods; (5) The Streator-Trilling Food Preparation Test; and (6) The King-Clark Food Test. It is worth noting that no one of these seven tests on the average as to validity meets the minimum desired correlation value, represented by a validity coefficient of 0.60, and that only one, Unstandardized Test C, meets the minimum reliability value desired, a coefficient level of 0.90. What is important is that the unstandardized tests do excel in comparison with standardized tests. The findings of the investigation are very similar to those previously reported for algebra tests and for silent reading tests.

A Study of Spelling Tests.—This investigation used five different spelling tests, two of them standardized instruments, and the other three teacher-made objective spelling examinations. The standardized tests were the spelling examination of The Progressive Achievement Tests, Elementary Battery, Forms A and B, and the Spelling Examination of the New Stanford Achievement Test, Advanced Examination, Forms Y and Z. In constructing the teacher-made, objective spelling tests, six lists of words equivalent in difficulty were chosen from the Buckingham Revision of the Ayers Spelling Scale and from the Iowa Spelling Scale. These words were then arranged in order of difficulty, so that for each of the six tests, 50 words increasing in difficulty were available. In the selection of the words, care was exercised to choose words well adapted to the types of tests planned for construction. The character of the unstandardized tests may be made clear by examples.

SPELLING TEST I

Directions: Below in each line you will find four words, only one of which is spelled correctly. In front of each word is a letter. Find the word that you think is spelled correctly. Write the letter in front of it on the blank at the beginning of each line. If you are not sure which word is spelled correctly, leave out the line, and continue with the next one. DO NOT GUESS!

Example: (b) 0. (a) sed (b) said (c) siad (d) saed

() 1. (a) forget (b) forgett (c) forgeet (d) forrget

() 50. (a) clerical (b) clerrical (c) clerical (d) clericule

Spelling Test II is a multiple response examination similar to Test I.

SPELLING TEST III

Directions: In each of the following sentences, the underlined word is spelled either correctly or incorrectly. If it is spelled correctly, place a check on the line before the letter R; if incorrectly, on the line before the letter W. If the word is misspelled, spell it correctly on the line at the right of the sentence including the underlined word. If you do not know which is the correct spelling of the word, skip the sentence, and continue with the next one. DO NOT GUESS!

Example:

 R ✓ W 0. Sixty seconds make one minit. 0. minute

 R W 1. Jack and Jill ran down the hill... 1.

 R W 50. That is no exageration..... 50.

Spelling Test IV is a combined true-false and recall test similar to Test III.

SPELLING TEST V

Directions: In this test, some letters in each word have been left out. Fill in the blanks with the letters that are not given, in order to finish spelling the word correctly. After each word is a definition or another word which means the same thing as the word that you are to spell. If you do not know how to spell the word, skip it, and go on with the next word. DO NOT GUESS!

Example: 0. anc__nt (very old).

1. S__th (a direction; the opposite of north).

50. gent__l (elegant in appearance; polite).

Test VI is a completion examination similar to Test V.

The data as to the reliability and the validity of these different tests are given in Table XIII.

The order of excellence, using as an index the median value of the validity coefficients, is as follows: (1) The New Stanford (Advanced Examination) Spelling Test (0.775); (2) The Progressive Achievement (Elementary Battery) Spelling Test (0.773); Tests III-IV, considered as equivalent forms of one examination (0.743); Tests V-VI, considered as equivalent forms of one

TABLE XIII.—MEASURES OF THE VALIDITY AND RELIABILITY OF CERTAIN SPELLING TESTS¹

Tests	Multiple Response II	True-False III	True-False IV	Completion V	Completion VI	Progressive Achievement A	Progressive Achievement B	New Stanford Z	New Stanford Y
Multiple Response I.....	0.804	0.743	0.697	0.680	0.585	0.575	0.694	0.592	0.666
II.....		0.741	0.692	0.692	0.577	0.670	0.729	0.622	0.698
True-False III.....			0.775	0.784	0.692	0.744	0.812	0.673	0.814
IV.....				0.756	0.723	0.791	0.782	0.737	0.739
Completion V.....					0.801	0.795	0.754	0.826	0.826
VI.....						0.658	0.721	0.821	0.755
Progressive Achievement A.....							0.933	0.795	0.869
B.....								0.812	0.865
New Stanford Z.....									0.858

NOTE.—The probable errors of the coefficients vary from 0.008 to 0.042; the number of cases is 109. The data were obtained in grade 6 of the National City, Calif., elementary schools during October and November, 1934.

¹ M. O. HIGER, A Comparative Study of Teacher-made, Objective Spelling Tests and Standardized Spelling Tests, p. 43, unpublished Master's thesis, University of Southern California, 1935.

examination (0.740); and Tests I-II, considered as equivalent forms of one examination (0.685). The standardized tests have an advantage in validity, but the margin is not large over two of the teacher-made objective spelling tests, and all five tests exceed markedly the desired minimum validity coefficient value of 0.60. The order of excellence as to reliability, in terms of the values of the reliability coefficients, is (1) The Progressive Achievement Spelling Test, (2) The New Stanford Spelling Test, (3) Tests I-II, (4) Tests V-VI, and (5) Tests III-IV. Only one of the

five tests exceeds the minimum desired reliability coefficient value of 0.90, but both standardized tests demonstrate an advantage in reliability over the three teacher-made objective spelling tests.

Perhaps it should be noted that proofreading skills are utilized in achievement in Spelling Tests I-IV, inclusive, and possibly

TABLE XIV.—A SUMMARY OF THE SCORES YIELDED BY FIVE GENERAL SCIENCE TESTS¹

Test	N	Range		M	σ	V ²
		From	To			
Ruch-Poponoe A.....	89	11	57	28.62	9.35	32.67
B.....	92	11	50	22.64	8.60	38.05
Dvorak R-1.....	92	71.5	103.5	88.20	6.65	7.54
S-2.....	86	70	95.5	81.88	5.45	6.66
Sones-Harry A.....	87	3	51	21.14	10.15	48.01
B.....	86	1	52	21.42	13.60	60.66
Teacher-made A.....	90	15	83	53.94	13.80	25.58
B.....	88	23	86	49.60	14.80	29.84

¹ J. D. CLARK, A Statistical Analysis of Teacher-made and Standardized Tests in General Science, p. 18, unpublished Master's thesis, University of Southern California, 1935.

² The coefficient of variation V is obtained by the following formula: $V = 100\sigma/M$. This means that with a normal distribution, since there are 5σ distances between 0 and the mean, V will be 20. V is the percentage of the distance between 0 and the mean occupied by 1σ distance.

also in Tests V-VI. This is not taught in school ordinarily, and the practice effects during the taking of the series of spelling tests may have interfered to lower the validity and the reliability coefficient values of the six forms of unstandardized spelling tests. The evidence indicates that classroom teachers can construct reliable and valid objective spelling tests, and the findings are in agreement, in general, with those previously reported for silent reading, algebra, and foods achievement tests.

A Study of General Science Tests.—The tests used in this study included the following measuring instruments:

The Sones-Harry High School Achievement Test, Science Examination, Forms A and B

The Dvorak General Science Test, Forms R-1 and S-2

The Ruch-Poponoe General Science Test, Forms A and B

Two teacher-made objective tests in general science, designated as Tests A and B

These tests were given to 94 pupils in grade 8 in a junior high school. Because of absence from school, not all of the pupils responded to all of the tests. Each form of the tests was given once only.

TABLE XV.—MEASURES OF THE VALIDITY AND RELIABILITY OF CERTAIN 9B GENERAL SCIENCE TESTS¹

Tests	Ruch-Poponoe B	Dvorak, R-1	Dvorak, S-2	Sones-Harry A	Sones-Harry B	Unstandardized A	Unstandardized B
Ruch-Poponoe A	0.832	0.636	0.632	0.851	0.794	0.696	0.698
B		0.680	0.725	0.764	0.559	0.652	0.694
Dvorak, R-1			0.628	0.696	0.652	0.632	0.524
S-2				0.708	0.621	0.674	0.682
Sones-Harry A					0.854	0.624	0.716
B						0.684	0.670
Unstandardized A							0.785

NOTE.—The probable errors of the coefficients vary from 0.02 to 0.06; the numbers vary from 86 to 92. The data were obtained in the Roosevelt Junior High School, San Diego, Calif., during the school year 1934-1935.

¹ CLARK, A Statistical Analysis of Teacher-made and Standardized Tests in General Science, p. 22.

Teacher-made Test A included 50 simple recall or single-answer completion items and 50 multiple-response (four-answer) items. Test B included 75 true-false items and 25 identification items. The content of these tests, arranged by the test constructor to be equivalent in difficulty on the basis of his own judgment, was checked against eight textbooks and workbooks commonly used in general science courses.

So far, no evidence has been presented in this chapter concerning the discrimination of the different types of examinations. The score series summaries obtained in this investigation are given in Table XIV.

The evidence in Table XIV indicates that the teacher-made objective general science tests are more discriminatory than any of the standardized general science tests. They are as nearly equivalent in total difficulty as are the standardized tests.

The data as to the reliability and validity of the test are given in Table XV.

The order of excellence, using as an index the median value of the validity coefficients for each test, is as follows: (1) The Ruch-Poponoe General Science Test (0.695); (2) The Sones-Harry High School Achievement Test, Science Examination, (0.690); (3) Teacher-made General Science Test B (0.688); (4) The Dvorak General Science Test S-2 (0.678); (5) Teacher-made General Science Test A (0.663); and (6) The Dvorak General Science Test R-1 (0.645). The order of excellence as to reliability, in terms of the reliability coefficients, is as follows: (1) The Sones-Harry High School Achievement Test, Science Examination; (2) The Ruch-Poponoe General Science Test; (3) Teacher-made Tests A-B, considered as equivalent forms of one examination; and (4) The Dvorak General Science Test. It should be noted that Form R-1 was constructed to be less difficult than Form S-2 of the Dvorak Test; the two forms are not equivalent, and therefore more evidence as to the reliability of the several tests is desirable. This evidence is presented in Table XVI.

TABLE XVI.—MEASURES OF RELIABILITY FOR SIX GENERAL SCIENCE TESTS¹

	<i>N</i>	<i>r</i> ± PE	PE _{meas}	PE _{meas} / <i>M</i>	PE _{meas} / <i>σ</i>
Ruch-Poponoe.....	89	0.832 ± 0.022	2.48	0.10	0.28
Dvorak R-1.....	92	0.695 ± 0.038	2.48	0.03	0.37
S-2.....	86	0.699 ± 0.036	2.02	0.02	0.37
Sones-Harry.....	86	0.854 ± 0.022	3.06	0.14	0.26
Teacher-made A.....	90	0.851 ± 0.020	3.59	0.07	0.26
B.....	88	0.849 ± 0.020	3.88	0.08	0.26

¹CLARK, *A Statistical Analysis of Teacher-made and Standardized Tests in General Science*, pp. 36, 39.

In Table XVI, the reliability coefficients for the two forms of the Dvorak Test and also for the two teacher-made objective tests were obtained by scoring the two halves of each individual form and then applying the Spearman-Brown formula.¹ Taking

¹G. M. RUCH and G. D. STODDARD, *Tests and Measurements in High School Instruction*, Pp. 356-361.

into account all of the measures of reliability given in Table XVI, the order of excellence, determined by ranks for each reliability measure and averaging the three ranks, is: (1) Teacher-made Test A; (2) Teacher-made Test B and The Sones-Harry High School Achievement Test, Science Examination, tied; (4) The Dvorak General Science Test, S-2; (5) The Ruch-Poponoe General Science Test; and (6) The Dvorak General Science Test, R-1. The reliability of all of these tests is, as Odell has described it, satisfactory for group measurement but only fairly so for individuals.¹ The evidence in this study is in agreement with that of the several investigations reported previously in this chapter; classroom teachers can construct general science tests that will be as valid and as reliable as standardized tests in the same subject and school grade.

The Advantages and Limitations of Standardized Tests.—Standardized tests have two advantages: (1) Ordinarily, they are much more carefully constructed than are teacher-made objective tests, and so fewer errors in measurement occur; and (2) all standardized tests possess norms. These advantages are closely associated with limitations, and both will be discussed.

Errors in measurement may occur because of improper or inadequate directions, because of errors in the scoring key, and because of the test content. Errors due to test content may be of two types: (1) those occurring because more than one correct answer is available or because no correct answer is available; and (2) constant errors. Constant errors may exist because the standardized tests include materials not stressed or, possibly, not even taught to the class being tested. Because standardized tests are intended for widespread use, they must include the content commonly taught in the given subject and school grade in many geographical areas. Some content is taught in different school grades in different geographical areas from others, and some is not taught at all in certain areas, whereas it is taught in other areas. The presence of subject matter that is widely but not universally taught is what causes constant errors of measurement for some pupils in schools or grades in which the given subject matter has not been taught. Obviously, unless teacher-made objective tests are constructed very carefully, more of the various sources of error will result from their use than from

¹ C. W. ODELL, *Educational Measurement in High School*, pp. 65.

the use of standardized tests. If sufficient care is given in the construction of teacher-made objective tests, errors can be minimized until there is little to choose between these two types of school examinations.

The one great advantage of standardized tests over teacher-made objective examinations lies in the norms that they possess. Norms are valuable for comparison against achievement standards, either of schools in general or of schools in a given area, such as a given state. To have this value, however, the norms must have been obtained through using the test experimentally in several school systems in several different geographical areas and with large numbers of cases. Norms resulting from using a standardized test in one school system only and with a few hundred pupils have little or no value for purposes of comparison. Such norms do not represent a cross section of pupil achievement. The average resulting from such an experimental use of the test might be too high or too low for safe comparison with results obtained with a class in any other area; unthinking acceptance of norms, without considering where and how they were obtained, and slavish efforts to meet norms with pupil average achievements constitute two serious errors on the part of teachers. In any case, even if the norms do represent a normal sampling of the total population, efforts just to meet the norm tend to mechanize education and educators. Such efforts tend to defeat progress, achieving only a static condition educationally. The effort should be to think of each individual's achievement in terms of the norms and in terms of his ability-to-learn and other characteristics. In so far as the group is concerned, the effort should be to exceed the norm, if this can be done without handicapping the pupils in any way.

The Advantages and Limitations of Teacher-made Objective Tests.—It has been noted that teacher-made objective examinations are subject to errors, resulting from carelessness in construction, to a greater extent than are standardized tests. On the other hand, teacher-made tests are less subject to the influence of constant errors although probably more subject to the influence of variable errors than are standardized tests. If the classroom teacher who is building an objective test will check each item included in the content of the test to make certain that it represents subject matter commonly taught in the given school grade

and subject and also that it represents material that has been presented carefully in her own instruction of the class to be tested, the data secured by the test will be little affected by such types of errors. It is possible, then, to construct teacher-made objective tests as error free as standardized tests.

Standardized tests do have the advantage of norms. The classroom teacher seldom uses an objective test with large numbers of pupils. It is possible to file such tests and to use them with several groups, with proper changes to avoid materials influenced by error, and to make comparisons between scores obtained by different groups of children. The advantage of standardized tests in this particular is partly overcome, in the writer's opinion, because the classroom teacher can make certain that no sources of constant error appear in the tests that she constructs.

Conclusions.—The evidence that has been submitted in this chapter clearly shows that Ruch's contention is correct. Classroom teachers can build teacher-made objective tests that have as good validity and reliability as the better standardized tests in the same subject and school grade. The evidence shows that this is not always done by every classroom teacher in constructing every objective examination. The general conclusion resulting from a study of the evidence is that education needs better standardized tests than are now available and more of them and also that classroom teachers should learn the fundamentals of test construction and of statistical analysis in order that they may construct better teacher-made objective examinations more generally than is now the case. Better measuring instruments, together with more intelligent analysis of data yielded by such instruments, will inevitably improve instruction.

In Chapter VI will be presented methods for constructing the more common types of teacher-made objective tests, together with an analysis of the advantages and limitations of each type of such examinations.

Questions and Exercises

1. Prepare a statement showing how teachers can take into account the majority of the characteristics of a good test (Chapter II, page 16) when building an objective test. Which characteristics cannot be met during construction? What can the teacher do about these characteristics?

2. How can a classroom teacher proceed to determine the reliability and the validity of her own tests?

3. Prepare a frequency table for the validity coefficients only in Table X, and then compute the mean and the standard deviation for each such distribution of coefficients. Interpret the results, with respect to the validity of the several tests.

4. What precautions should a classroom teacher take with respect to norms when selecting a standardized test for use with her class?

5. The data presented in Chapter V represent a variety of school grades and of subject fields. Why should this tend to permit one to determine on general conclusions concerning the problem under consideration?

CHAPTER VI

THE CONSTRUCTION AND USE OF TEACHER-MADE OBJECTIVE TESTS

The classroom teacher must be able to prepare and to use valid and reliable examinations of all types, both essay and objective tests, in order to analyze and to evaluate properly the products of pupil learning. Ordinarily, tests are used in schools for two general purposes, administrative and instructional. The school administration uses examinations to obtain school records of pupil achievements; for reports to parents concerning pupils; for purposes of promotion; for comparison of achievement of schools within the school system or with schools in other school systems; and for various research purposes, such as the evaluation of certain methods of instruction in comparison with other methods of instruction. Instructional tests are of two types: survey and diagnostic. The classroom teacher uses the survey type to determine, in general, how well a class has learned the subject matter that has been taught in a given period of time. Diagnostic tests are used to determine specifically with a single pupil or with an entire class what weakness exists following instruction and to provide information leading to remedial teaching designed to eradicate such weakness. Objective tests may be used for all of these purposes, in many instances more effectively than the traditional essay examinations. It follows, then, that the classroom teacher must know how to prepare and to use teacher-made objective tests if she is to maintain a high level of instructional efficiency.

The Characteristics of an Objective Test.—An objective test is an examination in which every response is either correct or incorrect from the standpoint of truth or fact. Theoretically, at least, correctness or incorrectness does not depend upon the personal opinion of the examiner. A good objective test permits no opportunity for disagreement among different examiners as to the correctness or incorrectness of individual responses to items in the test.

Objectivity may be secured in several ways. In the first place, directions should be provided to guide the student who is responding to the test. Such directions should be phrased in simple words, *i.e.*, words readily comprehensible to the students who are to respond to the test; they should be briefly but completely stated; and, ordinarily, they should include an example to illustrate just what the student is to do in responding.¹ In the second place, no items that are debatable as to truth or fact should be included in the content of the examination; *e.g.*, items as to the correctness of which experts disagree should not be used. In the third place, the phrasing of the items must be unambiguous. The statement of the item must be such that it is not only comprehensible to the pupil responding to the test but also allows of no disagreement among different examiners as to the correctness or incorrectness of the pupils' responses. Finally, a scoring key should be provided containing the correct answers, with only one answer correct for each item in the test.

For a complete evaluation of pupils' knowledge, objective tests should include more than one type of item.² Some pupils readily acquire the ability to respond to one type which may be quite difficult for other pupils in the same group. For example, excellence of memory plays some part in answering test items for which pupils must recall facts. If the same pupils were required only to identify the same facts in a collection of responses, a situation in which memory is of less importance to success, the achievement would be different from what it is when recall is necessary. For this reason, it is advisable to have more than one type of item included in an objective test.

Since the objective test forms permit rapidity of response, with many items answerable in a short period of time, it is advisable to limit the amount of writing that the pupil must do in answering any one item. When the written response is held to a minimum, objectivity of scoring is increased, since then it will not be necessary for the examiner to evaluate the correctness or incorrectness of partially correct answers. Furthermore, the necessity for

¹The use of an illustrative example is quite important when pupils are in the lower and middle elementary grades, particularly when they are not familiar with objective tests.

²The different types of objective test items are discussed in detail later in this chapter.

giving a complete and accurate answer in writing frequently interferes with the pupils' ability to give the correct information. The pupils become entangled in conflicting facts or involved in the mechanics of expression, so they give wrong facts or only part of the facts. Finally, an objective test should be so arranged that the pupils' responses are readily identifiable, both as to meaning and as to placement on the page. This makes for ease and speed of scoring. The matter will be discussed in detail later in this chapter.

A Simple Classification of Types of Objective Tests.—Many kinds of objective tests have been experimented with during the past two decades.¹ The more common types suffice for classroom purposes. The more commonly prepared types of objective teacher-made tests may be classified simply as follows:

- I. Recall types:
 1. Simple recall items (single answer per item)
 2. Completion exercises
- II. Recognition types:
 1. True-false exercises (the true-false-neither examination is actually a three-answer multiple-response examination)
 2. Multiple-response exercises, including the so-called best answer, or judgment, test
 3. Matching exercises
 4. Identification exercises
 5. Rearrangement exercises

Combinations and variants of these types of tests are quite commonly found, as well as other types of objective examinations.

Advantages of Objective Tests.—Certain general advantages pertain to teacher-made objective tests. These will be discussed briefly. The first great advantage is the objectivity of the test; this reduces the influence of the examiner in determining achievement to the status of a negligible factor. It insures the recognition of achievement to those to whom recognition is due; total scores cannot be influenced by the personal bias of the examiner or by halo effects, negative or positive. Tests can be made objective if sufficient care is used in their construction.

¹ For specimen examinations, see H. D. RINSLAND, *Constructing Tests and Grading in Elementary and High School Subjects*, New York: Prentice-Hall, Inc., 1937.

G. M. RUCH and G. D. RICE, *Specimen Objective Examinations*.

Another advantage is that objective tests permit a wide sampling of students' knowledge in quite short periods of time. It is true that the sampling is extensive rather than intensive, as is frequently the case when the traditional, or essay, examination is used, but extensiveness is of value when time is limited.¹ The objectivity of such tests, coupled with extensive sampling, other things being equal, tends to insure valid and reliable examinations. If the items selected for an objective test are those commonly stressed in good teaching practice in the given grade and school subject, and if the test is objective and of considerable length, it will be both valid and reliable to a degree greater than any essay examination that can be given in the same time.² In the construction of objective tests, it is possible to arrange the items on the test pages so that scoring is easy and rapid. This means that time in preparing the test is substituted for time in scoring the completed examinations. Also, scoring an essay examination involves careful, continuous evaluation of the students' reactions, and this requires alertness and considerable mental effort. Scoring a properly devised objective test does not involve great mental labor. These matters will be discussed further hereafter in connection with the several types of objective tests.

Objective tests have definite advantages over essay examinations for improving instruction. Pupils responding to essay examinations quite frequently use different facts in their discussion of a given test item, and they seldom give all of the pertinent facts completely and accurately. It is common for extraneous

¹ The essay, or discussion, type of examination is well known to teachers. It seems necessary only to illustrate rather than to discuss it extensively. A typical test, devised for use with upper grade pupils in a traffic safety course, follows:

1. What is meant by highway courtesy and discourtesy?
2. How are habits in automobile driving formed?
3. Why has it been necessary to insist on drivers being licensed? What kinds of drivers' licenses are now in use?
4. Have accidents decreased in Texas since drivers' licenses have been issued? Give definite facts in answering this question.
5. What are some of the ways in which drivers can avoid accidents?

² H. D. RINSLAND, *Constructing Tests and Grading in Elementary and High School Subjects*, pp. 295-299.

G. M. RUCH, *The Objective or New-type Examination*, pp. 89-100, 281-306.

material to appear in answers to essay test questions. For these reasons, it is impossible to prepare any detailed analysis of such an examination tending to show conclusively what and how much remedial instruction is necessary. Objective tests, on the other hand, readily permit such an analysis, utilizing the error analysis technique described in Chapter IV (page 72). When the error analysis chart has been considered, it is possible to diagnose accurately what remedial instruction is necessary to insure mastery of the subject matter by all of the pupils in the given group.

The objective test permits the teacher definitely to control the examination. The pupil is compelled to respond to subject-matter situations that the teacher knows are important; he cannot bluff, distorting the question to make it fit whatever knowledge he happens to have retained. Furthermore, the objective test is economical of pupil effort; writing is reduced to a minimum, so the pupil is free to concentrate his thinking on the subject matter of the test rather than on his responses to the test items. Also, since the teacher can construct duplicate or almost duplicate tests readily, it is possible to control testing at more than one time.

Finally, informal objective tests are more valuable in measuring pupil achievement for small units of subject matter than are standardized tests or essay examinations. Standardized tests must be devised for wide usage, and frequently such tests contain subject matter not stressed in a local situation. Essay tests give relatively small coverage of what has been taught; the time necessary for organizing information and for stating the response to the test item makes it impossible to respond to many individual test items. Objective tests, on the other hand, permit a wide coverage in relatively little time.

It should not be thought that objective tests are free from handicaps, even though they are superior in many respects to other types of examinations for many purposes of measurement. It is necessary now to consider the limitations affecting objective examinations.

Limitations of Objective Tests.—Objective tests deal largely with information presented very briefly. Because of this, they may become measures of factual information only, concentrating upon the recall or recognition of facts and avoiding demands on

reasoning or on judgment. Since recognition types of tests are greater in number than recall types, pupils may need to retain few of the acquisitions of the classroom in a form capable for recall and for use in relationship to other facts.

This suggests a second serious limitation of objective tests. Because objective examinations require only brief responses, they provide no opportunity for the organization of thoughts or for the expression of thought organizations in grammatically correct sentences, paragraphs, and groups of paragraphs. Yet it is difficult to conceive of much beneficial training resulting among students during the relatively brief periods of time devoted to essay examinations, particularly since the pupils must select rapidly the pertinent facts from a mass of information, organize these facts into a meaningful whole, and then write rapidly under pressure. The difficulty is that the student responding to an essay test may not give all of the information accurately and completely, and there is always the danger of mechanical errors or of improper thought organization. Since no time is allotted for correcting and recopying an essay examination response, any training in the use of English may be detrimental rather than beneficial. Slow writers, too, are penalized on essay tests. Finally, teachers tend to confuse pupils' ability to use good English and to express their thinking in writing with their knowledge of the subject matter of the examination. The point to be brought out is this: Objective tests suffer from this limitation, but they at least do not offer training in incorrect or fragmentary responses, as do essay examinations. In any case, training in the use of good English and in the proper and adequate organization of thought should be given in classrooms otherwise than by examinations. An examination is a measuring instrument that samples more or less adequately the kinds of learning accomplished in school classrooms and the degrees of subject-matter mastery accomplished by pupils during the learning.

A third limitation arises from the possibility of using statistical procedures and techniques in connection with the results of objective tests. Often the temptation to analyze test total scores statistically, particularly with reference to the establishing of school marks, causes a teacher to lose sight of the diagnostic values of an objective test, and so the instructional advantages of the test are disregarded.

A fourth limitation lies in the guessing effects on certain types of tests, disturbing the accuracy of the results to a greater or lesser degree. It has been urged that recognition types of tests promote guessing more than do essay examinations. This is not true, however, if a sufficient number of alternative responses are available. It is true that certain types of objective tests do tend to confuse pupils by presenting false statements, incorrect answers, and similar material, but this confusion will not exist if the subject matter has been well learned and *if the tests are used as teaching material after they have been scored*. If pupils are left to ascertain their own errors, confusion is likely to result, and this probably is more true of younger pupils than of older pupils.

A fifth limitation of objective tests is their cost. They usually require some form of duplication, so that every pupil may have an individual copy. This difficulty may be overcome in part by the use of inexpensive duplicating devices, such as the hectograph, or by using the same mimeographed test papers several different times, inexpensive scoring sheets being utilized for the pupils' responses. Furthermore, certain types of tests may be dictated, and others may be printed on large sheets of inexpensive paper, with large-size type visible throughout the classroom. Such printed tests may be used several times for testing and for instruction after testing. A secondary limitation is that recurrent uses of tests are possible only after the tests have been refined by tryouts of the materials in actual tests, but even so the cost per test per pupil is materially decreased by this procedure.

Finally, it has been urged that objective tests are unnatural and unpedagogical, that information is never isolated in actual usage in the manner in which it is isolated in objective test items. The same thought applies to almost any other type of examination that might be used in classrooms. As long as pupils are asked to achieve in schools, some form of accounting of results is necessary, and objective tests function at least as adequately for this purpose as do other kinds of examinations. The questions might be asked: What is natural and pedagogical? Are only old and familiar forms of tests natural and pedagogical?

With these general advantages and limitations of objective tests in mind, it is possible to turn to a consideration of the more common types of objective tests in use in classrooms. Among the matters to be considered are the form of such tests and their

advantages and limitations. Matters pertaining to the construction of objective tests will be considered briefly, with a more detailed presentation to be given later.

Types of Tests. *Simple Recall and Completion Tests.*—Because of their similarity, these two types of tests are considered together. In effect, the simple recall examination is only a one-blank-per-item completion test. The completion type of test is the most often used form of objective examination.¹ Ordinarily, simple recall tests should be used in lower elementary grades, whereas both sentence completion and simple recall tests may be used successfully in middle and upper elementary grades.

There are a number of matters to be considered when one is preparing a simple recall or a sentence completion examination. The scope and the length of the test must be determined before the test construction is begun. After this, the directions for the pupil must be prepared. These must be expressed in simple language, clearly phrased, and in sufficient detail so that the pupil cannot be misled. Usually, in the elementary grades, it is desirable to include an example in the directions to the pupils.

After this has been accomplished, the test maker is ready to prepare the content of the test. The test content should include subject matter commonly stressed in good teaching practice in the given grade and school subject, preferably that which has been emphasized during the teaching preceding the giving of the test. It is good practice to select the content for a test during the period of instruction. It is not enough, as the period of teaching draws to a conclusion, merely to extract statements from one or more textbooks, deleting occasional words from each statement selected. The test maker should construct definite sentence statements with a central fact or word or date omitted. This avoids the use of responses that can be answered by the pupil from his general information, possibly without any great amount of specific knowledge of the subject matter of the test. Statements should be definite, demanding a single short answer whenever possible. Unless the item statement is carefully phrased to demand a single specific answer, optional answers as to dates and places may occur. This is particularly true in the social sciences.

¹ G. M. RUCH, *The Objective or New-type Examination*, p. 189.

It is difficult, if not impossible, to prepare a simple recall or a completion test for which numerous alternate and correct answers may be given. Although each blank should call for but a single idea, synonyms, different facts, and the like may appear correctly in the test. It is essential that all correct answers appear in the test key for all test items. The presence of numerous equivalent answers in the key, all equally credited, does not mean that the scoring is subjective.

The number of blanks affects the difficulty of the test; ordinarily, the more blanks in a sentence completion item the more difficult is the item. This is the reason that the simple recall type of test is preferable for lower elementary-grade pupils. The test items should not be puzzle situations to pupils; ordinarily, two or three blanks in a sentence completion statement are sufficient for upper elementary-grade pupils. These blanks may appear anywhere in the statement. In the simple recall test, blanks should be terminal and aligned. They should appear at the right of the test page, conforming to normal reading habits of pupils.

The test maker should avoid using words that tend to make the correct responses obvious. Cue words, such as *a* and *an*, occurring immediately before a blank in the test item sometimes indicate the character of the desired response. If this precaution conflicts at times with the necessity for securing definiteness of statement, the latter takes precedence.

HEALTH TEST 2

Your name _____ Date _____ Score _____

Directions: Below are a number of sentences. Each needs one word to make it a true statement. This word is the last in each of the sentences. Write the missing word on the line at the right of each sentence.

Example:

0. The body as a whole benefits from (0) exercise

1. Exercise helps the body to regulate its. (1) _____

2. The lungs are enlarged by. (2) _____

30. Muscles are developed to do their work well by (30) _____

When all the test content has been prepared, carefully checked, and arranged by the test maker roughly in order of difficulty, there remains the task of preparing directions for the examiner.

These should indicate clearly and concisely what the examiner does as to distributing the tests, checking to see that pupils have pencils, and the like; the time allowed for the taking of the test; the scoring procedure; and the amount of credit allowed for each correct response, one point of credit per correctly answered blank. Ordinarily, it is not necessary to include instructions as to the statistical handling of the data yielded by the examination.

An example of the simple recall test is shown on page 113.

This type of test should be scored with a strip stencil made of colored paper at least 3 inches in width. If the scoring is to be done by artificial light, the paper should not be glazed; and it should be relatively light in color. The difference in color between the test paper and the stencil, plus the width of the stencil, makes it easy to score quickly and without strain. The width makes it easy for the scorer to concentrate on two things for each item: (1) the correct answer and (2) the pupil's answer. Since the correct answers should be placed at the right of the stencil, these two things will be in close juxtaposition. The first and the last answers on each stencil strip should be numbered and underlined, to facilitate the matching of the key answers with the pupil's answers by items. Using the sample partial test, the scoring stencil in position for checking the pupil's responses is shown below.

Your name _____ Date _____ Score _____

Directions: Below are a number of sentences. Each one of these sentences needs one word to make it a true statement. This word is the last word in each of the sentences. Write the missing word on the line at the right of each sentence.

Example:

0. The body as a whole benefits from (0) _____
exercise

1. Exercise helps the body to re _____
1. temperature _____
2. The lungs are enlarged by. _____
2. exercise _____

30. Muscles are developed to do _____
30. exercise _____

An example of the simple recall test prepared for use in upper elementary grades follows.

TEST ON THE EXPLORATION AND COLONIZATION OF TEXAS¹

Your name _____ Date _____ Score _____

Directions: Write the correct answer to each question below in the blank following the question at the right. Work rapidly and accurately.

Example: In what year did Columbus discover America? . . . 1492 _____

1. Who was the first Frenchman to explore Texas? . . . _____
2. Who were the bravest and most admirable Indians in Texas? _____
49. What are the names of four missions near San Antonio? . . . _____
50. To what Mexican state was Texas joined? . . . _____

The scoring of this test is similar to that of the first one shown. The difference in form between the two partial tests is this: The first employs the partial-statement form of item; whereas the second employs the question form. The first was devised for use in lower grades; the second, for use in upper grades.

A sample partial-completion test devised for use in upper elementary grades follows.²

GEOGRAPHY OF THE UNITED STATES

Your name _____ Grade _____ Date _____ Score _____

Directions: In each of the sentences below, fill in the blanks with the correct answers. Work rapidly.

Example: 0. The greatest waterfall in the United States is Niagara Falls.

1. A great highland in the eastern part of the United States is the _____ highland.
2. Philadelphia is the _____ largest city in the United States. It is noted for the manufacture of _____.
59. _____ leads the United States in the production of sugar cane.
60. The two chief grains of the North Central States are _____ and _____.

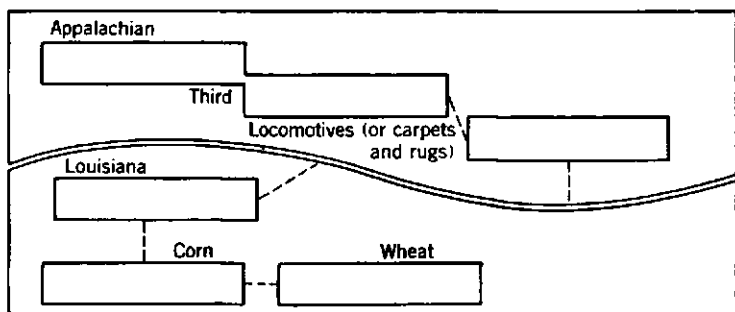
The easiest method for scoring a sentence completion test is to use a window stencil. Certain writers suggest the use of a double blank, one in the test item that is numbered; and a second, similarly numbered, at the right side of the pages and on the same line as the numbered blank. This makes possible the use of a strip stencil, but it is an unnatural form for pupils to use. Other

¹ Devised by Dorothy Emig, Houston School, El Paso, Tex.

² Devised by Beatrice Sims, Dudley School, El Paso, Tex.

writers have suggested the use of transparent stencils, but these are difficult to use, since it is impossible to mark errors on the papers, and also because such stencils tend to reflect light into the eyes of the scorer.

To make a window stencil, take an unused copy of the test, and cut an oblong around each blank. Lay this copy over a blank sheet of paper, and trace all of the windows (the oblongs) on the blank sheet. Then cut the tracings, border each one in black, and put in guide lines from one window to the next. This insures the scorer's checking each pupil response on the test paper. Write the correct response that should appear in each window on the stencil, preferably at the left of the window, since the eye reads from left to right. When all responses on the key have been placed on the stencil, it is ready for use in scoring. The window stencil for use with the example partial test given on page 115 appears below.



Finally, it seems appropriate to consider briefly the advantages and the limitations that affect simple recall and sentence completion tests. The first advantage of this type of test is its naturalness in mechanical form; the test items parallel typical question or statement forms. Recall types of tests are relatively free from the effects of guessing or of chance; they may be made almost entirely objective, particularly when all correct answers are given in the key, if more than one alternate answer is correct. This is particularly true of the simple recall test. Finally, they are easily scored, particularly the simple recall test.

On the other hand, recall types of tests tend to be highly factual in character, as one method for avoiding subjectivity in scoring. This suggests a second limitation: Unless care is exercised in

preparing the scoring key for the simple recall or completion test, scoring may become subjective. Also, the use of too many blanks in one item or the use of blanks for relatively insignificant words or facts tends to reduce the particular exercise to a puzzle situation. Finally, recall types of tests require more time both in administration and in scoring than recognition types. (Certain exceptions which will be noted later.) This criticism is applicable more to the sentence completion test than to the simple recall examination.

The types of tests with which the previous discussion has dealt require the memory of specific information. There are types of objective tests that require the recognition of correct facts from numbers of facts presented, some of which are correct, and some, possibly, not correct. These types of tests are called recognition tests, and the following discussion is concerned with the more common types.

Recognition Tests. *True-false Examinations.*—Actually, the true-false test is only one form of alternative response examination, but it has been so widely used and it differs enough in form and scoring from other forms of alternative response examination to warrant a separate treatment. True-false tests have been used almost as much as have completion tests.¹ They can be used in many school subjects and in all elementary-school grades in which objective tests are of value in measuring.

Certain considerations must be kept in mind when the test maker is devising a true-false test. All statements included in the test items definitely should be either true or false. Each statement should be expressed always in as few words as possible, using the language of the group responding to the test and correct mechanics of expression. Ambiguous and misleading expressions, such as trick or puzzle items, should be avoided. Similarly, long, involved statements and statements including double negatives should be avoided in elementary school tests, since comprehension of such items, necessarily a preliminary to correctness of response, is more important in attaining success than is information. Each item should be independent of all other items in the test; no response to any item should be based on an answer to any other. If possible, the pupils' responses should include no writing. This matter will be discussed in more detail here-

¹ G. M. RUCH, *The Objective or New-type Examination*, p. 189.

after, in connection with form of test presentation. These several considerations are helpful in general only; but a careful study of this paragraph should convince the test maker that preparing a true-false test implies much more than merely selecting a number of short statements from one or more textbooks and arranging them, with directions, in a test in which approximately one-half of the items are true and the remainder false.

As with simple recall and sentence completion tests, the test maker must prepare the test items; arrange them approximately in the order of difficulty; and prepare the directions to the pupils, including an example, and the directions to the examiner. The directions to the examiner cover the giving of the test and the scoring of the pupils' responses.¹

A sample true-false test devised for use in grade three follows:

TEST ON MILK AND THE MILK INDUSTRY

This tells you what you are to do. First write your name on this line:
Name _____

Now read the directions which tell you what to do.

Directions: Some of the statements in this test are true, but some are not true. Read each sentence carefully. If you think it is incorrect or false, draw a line under the word FALSE before it. If you think it is correct or true, draw a line under the word TRUE. The sentence below is a sample.

TRUE FALSE 0. Cows have four legs.

Cows do have four legs, so the statement is true, and the word TRUE is underlined. If you finish the test before the period is over, look over your paper, and correct any mistakes you have made. If you do not know, do NOT guess.

TRUE FALSE 1. Butter is made from skimmed milk.

TRUE FALSE 2. Milk is chilled to kill germs.

TRUE FALSE 30. Sour milk is iced to separate the curds from the whey during the making of cottage cheese.

Some writers advocate the use of different methods in recording the pupils' responses. Some of these methods will be demonstrated, using sentence 1 from the partial-sample test above. Since this item is false, only the indication for a false statement will be shown, with the symbol indicating true given apart from the item.

¹ These matters will be omitted hereafter in the discussion of the remaining types of recognition tests.

— 1. Butter is made from skimmed milk. If this statement were true, the symbol + would appear on the line at the left. Sometimes + and 0 are used instead of + and - to indicate true and false statements. Other test makers prefer that pupils use T and F or Y and N to indicate true and false statements. Because it is not difficult to change a - to a + and a capital T to a capital F (either written or printed), the use of these symbols should be avoided. Another example, using the same item, follows:

T ⊕ 1. Butter is made from skimmed milk. In this instance, the pupil is asked to circle the letter T if the statement is true; the letter F, if the statement is false. This is much superior to the methods used in the previous example. If the student circles the wrong symbol, he needs only to place a large X through the symbol and its circle and then draw a circle around the correct answer. A third example, still using the same item, follows:

1. Butter is made from skimmed milk. The pupil, in this instance, is told to place a circle around the number of the sentence if the statement is true and to make no mark on the paper for the item if the statement is false. This method is bad, since it forces the pupil to guess on all items concerning which he does not know. He should have the option of avoiding an answer if and when he wishes to do so. Failing to answer an item means that he has enforced a penalty against himself, since a correct answer would earn credit toward a higher final score than results when one or more items are avoided.

Probably the best method of response for lower grade pupils has been demonstrated in the sample partial test. The best method with upper grade tests appears in a second sample partial test.

A TEST ON AVIATION

Your name _____ Grade _____ Score _____

Directions: Some of the statements below are true, and some are false. If a statement is true, make a check on the line before the T at the left of the sentence number; if false, before the F. If you are in doubt, omit the item, and go on at once with the next statement.

Example: _____ T F 0. An airplane is a lighter than air flying machine.

- _____ T _____ F 1. A super charger is a device for compressing the air supply of aircraft engines.
- _____ T _____ F 2. An airplane designed to rise from and alight on water is called an amphibian.
- _____ T _____ F 50. Minimum speed is the lowest speed of an airplane at which control can be maintained.

This type of response situation calls for a minimum of pupil writing in giving a response to the test. If a wrong response is given, it is possible to correct it merely by encircling the incorrect check and checking the correct line. Furthermore, this type of test can be quickly scored by using a strip stencil, cut three typewriter spaces wide, which is placed over the T and between the two lines for the responses. The correct response is indicated by a small ink line on the proper side of the strip. In this instance, scoring becomes as rapid as counting. The scoring of other types of responses is slower and more difficult.

Any discussion of true-false tests is incomplete without mention of dictated true-false examinations. Such tests sometimes are used by instructors as a means to economy. Statements in dictated true-false tests must be short, preferably no longer than those in the sample partial third-grade test presented heretofore (page 118). The human memory cannot cope quickly with long, involved statements, and their use reduces a dictated true-false test to a guessing contest. Some pupils will respond better to dictated examinations than to printed tests, but this is not typical of the majority of individuals. Ordinarily, the use of dictated true-false tests is to be avoided, particularly in lower and middle elementary grades.

One variant of the true-false test is the true-false-neither examination. This is set up in much the same form as the second sample partial test given above (page 119), except that the directions provide for the third response, and the column for the pupils' responses appears as follows: _____ T _____ F _____ N, etc. This test type should never be used below grade 6, as it is too difficult for middle and lower grade pupils. It has certain advantages, in that it permits the use of items concerning the truth or falsity of which authorities disagree, of items including a partially true or partially false statement or both, and of items that are true but do not tell all the truth. This form of test requires

that pupils responding to it weigh each statement carefully, in every instance exercising judgment in arriving at their response. This changes what is otherwise largely a factual test into a reasoning examination. The T-F-N test is much more difficult to construct and to score than is the ordinary true-false test, and it requires much more time for pupil responses than does the ordinary true-false test.

Some test makers attempt to minimize scoring difficulties by arranging test items for pattern responses. For example, it is possible to arrange items in alternate order, the first one true; the second, false; and so on. Another method is to arrange items by fives, the first two true, and the next three false, reversing this order in the second group of five items and continuing in this manner throughout the test. This is objectionable, since alert pupils readily note the pattern and follow it rather than replying to the items on the basis of their information.

Ordinarily, one-half the items in any true-false test should be true; and the remainder, false. When the T-F-N form is used, the arrangement should be one-third true, one-third false, and one-third neither.

The formula for arriving at a total score on a true-false test is to subtract the false from the true answers ($R-W$). Omitted responses should not be counted as errors; so to count them involves placing a double penalty on the pupil. The formula for the T-F-N test is: $\text{Score} = R - \frac{1}{2}W$. Total scores obtained in this way agree almost exactly with actual total difficulty scores determined for the same pupil responses to the given examination.¹ The correction formula eliminates almost entirely the effects of chance guessing on the part of pupils in determining their responses.

True-false tests have certain advantages and also certain limitations. They may be used in many school subjects and grades; and they obtain an extensive sampling in a minimum of time. On no other objective test is it possible to obtain as many individual item responses in a given time as on the true-false test. This type is easier to construct than many other types, although the difficulties of the test maker in securing defensible items are greater than is readily apparent. Once the scoring

¹ For example, see M. E. BROOM and L. H. LOVELACE, *A Study of a Test in Home Mechanics*, *High School Teacher*, 3: 262-263, 272 (1927).

key has been established, scoring is perfectly objective. The judgment of the test maker in determining the truth or the falsity of an item makes scoring objective thereafter. The T-F-N form is an excellent reasoning test, allowing the use of much subject matter not otherwise available for use in this type of examination.

True-false tests, unless they are used as teaching instruments after their use as measuring instruments, tend to confuse pupils and even to teach them erroneous facts. Much important subject matter is neither absolutely true nor absolutely false, and some test makers use such material subjectively indicating it to be true or false (more true than false, and vice versa). Occasionally, this causes scoring difficulties, since some examiners alter the key according to their subjective judgment, maintaining that any item that is only partially true should be considered entirely false, and so on. Unless the correction formula ($R - W$) is applied in determining total scores yielded by a true-false test, the scores will be inaccurate, since they will be affected by chance guessing. Unless statements have been carefully prepared, the presence of ambiguous words and of negatives tends to confuse the meaning of the items. Most of these limitations can be avoided by careful test construction and proper use of the tests, including correct scoring.

Possibly enough has been presented concerning the first of the alternative types of examinations, the true-false test. The following discussion will be concerned with the multiple-response examination.

The Multiple-response Test.—This type of examination includes items in each of which a statement or question is given with two or more alternate answers, only one of which is correct. When this correct response is indicated by the pupil responding to the test, the item makes a complete statement that is correct in its information.

Each item in a multiple-response test should have the same number of alternate responses, so chosen that all such responses possess plausibility. Cue words should be avoided, and each alternate response should have the same grammatical form as the main portion of the item, so that each of the responses makes complete sense when read alone with the item statement. The correct answers should be distributed at random, with about the

A TEST ON TEXTILE MANUFACTURING IN THE UNITED STATES

Directions: In each of the exercises below, you will find an incomplete sentence followed by four words or phrases, one of which will complete the sentence correctly. Each of the four words or phrases is numbered. You are to write the number of the correct answer on the short line at the left of the sentence. Look at the example below. Do your work rapidly.

Example: 1 0. The invention that made cotton growing profitable was the

1. cotton gin 2. compress 3. tractor 4. motor truck

_____ 1. The state that ranks highest in cotton manufacturing is

1. North Carolina 2. Texas 3. Nevada 4. South Carolina

_____ 2. Linen is made from

1. hemp 2. flax 3. cotton 4. wool

_____ 40. Textile manufacturing is the making of

1. paper 2. cloth 3. lumber 4. machinery

The advantage of this form of presentation lies in its ease of scoring. If a strip stencil 3 inches wide is used, the numbers of the correct responses being near the left-hand margin with the first and last response for each page on the stencil underlined, scoring is very quickly accomplished. This is probably the most rapid scoring technique known.

The sample test illustrated above is essentially a factual one. The multiple-response form is adoptable to reasoning examinations also. When used in this way, the arrangement of the test materials is similar to that in the example shown, but the alternate answers are relatively long statements, and more than one response is either entirely or partially correct. The pupil is instructed to indicate the *best* answer among those given. The alternate answers are arranged vertically one below another, instead of horizontally side by side, as shown in the sample partial test illustrating the multiple-response form. This requires more space for presenting the test materials, increasing the cost of the test per pupil. This fact, plus the difficulty of this type of test, restricts its use in the elementary school to the upper grade levels.

When the test content has been carefully checked to insure only one correct answer per item, scoring is entirely objective. Also, the multiple-response examination is very rapidly scored, par-

ticularly when the responses are numbered and the pupil indicates his answer by number. Multiple-response tests are easier to prepare than certain other types of tests, particularly completion, identification, and matching examinations. Finally, guessing can be minimized (eliminated, for all practical purposes) by using from four to seven alternate responses. One great advantage of this type of test, quite apart from those given above, is its availability for use as a reasoning test. Very few other objective types are adaptable to this use.

The last named advantage suggests one of the weaknesses of the multiple-response examination; ordinarily, it tends to be highly factual. In the factual type particularly, the incorrect answers frequently tend to confuse the pupil and even to teach him erroneous facts. The only remedy for this is to use the test as a teaching instrument after it has been scored.

Multiple-response tests are more or less open to chance guessing unless four or more alternate answers are given, even though the correct responses do not appear in the same position or order in successive items. It is often difficult to secure three plausible but incorrect answers, although it is true that to an uninformed pupil almost any answer will appear plausible. The same number of alternate responses must appear in all items, to minimize guessing and to facilitate the use of a correction formula, if necessary, so that much subject matter is not available for use in this type of examination. If less than four alternate answers are used, the correction formulas described in connection with scoring procedures for true-false tests must be applied in determining final scores.

There is always some danger of having more than one correct or nearly correct answer among the alternate responses for a given item. This, of course, is an advantage in the best answer multiple-response test, but it is definitely a limitation in the factual-type examination. The lengthy-statement best answer test requires much space, and it is difficult and relatively expensive to prepare. Because of space requirements, if used for reasonable purposes, the number of responses must be kept from three to five, thus allowing for chance guessing in some instances. The correction for this is the use of a scoring formula.

Among the less common but still frequently used types of recognition tests are the matching test, the identification test, and

the rearrangement test. They will now be considered in the order of mention above.

The Matching Test.—This is a form of alternate response examination in which the pupil responding to the test indicates which single item in one of two given columns is paired with one item in the second column. Words, dates, sentences, and the like may be matched. The matching test may be used also as a variant of the sentence completion test, matching blanks in sentences in one column with answers in a second column. This type of test is relatively difficult to construct.

Relatively little subject matter lends itself readily to use in matching tests. Only items that are absolutely definite and correct are available for use in this type of examination. Similar subject matter should be used in matching tests; for example, if one date is included in such a test, numerous dates should appear. Unimportant words or facts should not be included in this type of test, and cue words, such as one singular word in a group of plural words should be avoided always.

A matching exercise ordinarily should include 10 items in the left-hand column. Fewer than 10 promotes chance guessing, and more is uneconomical of time. The right-hand column should include at least 12 items, to assist in reducing success by chance guessing during the matching of the final few pairs in the exercise. If more than 10 pairs are desired in the matching test, the items should be grouped by exercises or divisions of 10 items in the left-hand column, and the directions should specify that the pupil responding to the test complete first the items in division 1, then the items in division 2, and so on. It has been suggested that the items in the right-hand column be arranged in alphabetical order.¹

The correct form for column matching is shown in the sample partial test below.

TEST ON THE EXPLORATION AND COLONIZATION OF TEXAS²

Directions: In the left-hand column below are 10 incomplete statements. You are to place the number of each statement from the left-hand column on the line before the statement in the right-hand column which completes the sentence correctly. Look at the example below:

¹ H. D. RINSLAND, *Constructing Tests and Grading in Elementary and High School Subjects*, p. 101.

² Devised by Dorothy Emig, Houston School, El Paso, Tex.

9. Columbus discovered America in

_____ 1492.

Since the date, 1492, in the right-hand column correctly completes the sentence in the left-hand column, the number 0 should be placed on the left at the left of the date 1492. The answers in the right-hand column are mixed up, and there are more of them than there are statements in the left-hand column, so do not expect to have a number on each line in the right-hand column. Work rapidly.

- | | | |
|---|-------|-------------------|
| 1. The Spanish Missions were established | _____ | were those near |
| | | San Antonio |
| 2. La Salle's Fort was named | _____ | to find gold. |
| | | |
| 10. The man who drew the first map of Texas | _____ | was San Felipe. |
| | _____ | to hunt and trap. |
| | _____ | Lavaca Bay |

The matching test is completely objective; and when facts only are involved, it is relatively easy to construct. The matching-test form may be developed as a judgment or reasoning examination, using long statements dealing with closely related subject matter, but it is difficult to construct, and it also is very difficult for even upper grade elementary-school pupils. Matching tests are very easily and rapidly scored, using a strip stencil.

Unless the precautions outlined previously are taken, matching tests suffer from chance guessing, when the number of items to be matched are too few, or from loss of time, when the number of items are too many. As with other alternate response examination types, incorrect answers tend to confuse the pupil and even to teach him erroneous facts. In certain situations, as in situations involving chronological dates very closely related in time, presented in a history test, the degree of discrimination required frequently cannot be defended against attack on the grounds that the exercise is a puzzle situation only.

The Identification Test.—Essentially this type of recognition test is a form of the matching test. Pictures are used in place of the left-hand column, and the pupil responding identifies facts in a column, using numbers on the picture instead of the numbers of facts, words, or statements in the left-hand column of a matching test. One format of this type of examination uses the completion form, listing the numbers in the picture in a column, each with a following blank. In responding to this form of identification test, the pupil supplies the word or words necessary in each

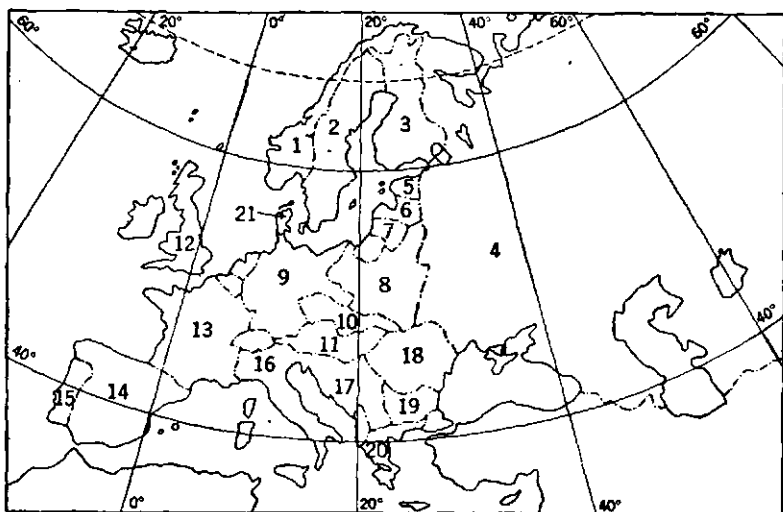
blank. Obviously, the cautions and procedures applicable to matching and completion tests apply to this type of examination.

A sample test follows.¹

NOTE.—This test was constructed and used during 1937–1938.

TEST ON EUROPEAN GEOGRAPHY

Directions: Below you will find pictures and lists of words at the right of each picture. The first picture is an outline map of Europe, with 21 countries numbered. Denmark is No. 21; and you will find the number 21 on the line at the left of the word Denmark in the list of countries at the right of the map. Place on the line before the name of each country the number that you find within the borders of that country on the map. There are more countries named in the list than there are numbered on the map, so do not expect to have a number on the line before each country named in the list. Do the following exercises similarly. Work rapidly.



_____ Belgium	_____ Great Britain	_____ Portugal
_____ Bulgaria	_____ Greece	_____ Rumania
_____ Czecho-Slovakia	_____ Hungary	_____ Russia
21 _____ Denmark	_____ Italy	_____ Spain
_____ Estonia	_____ Latvia	_____ Sweden
_____ Finland	_____ Lithuania	_____ Switzerland
_____ France	_____ Norway	_____ Yugoslavia
_____ Germany	_____ Poland	

¹ Devised by Catherine McCarthy, Houston School, El Paso, Tex.

The identification test is adequate for the measurement of factual information, particularly in science, geography, and similar subject matter. It can be devised to allow the testing of ability to apply principles to concrete situations or to identify examples of such applications to principles. On the other hand, it is limited to relatively few subjects in the curriculum; it is difficult to prepare, and financially expensive in its preparation, since the use of drawings requires much space; and incorrect answers tend to confuse pupils subsequently, as with all recognition-type tests, unless the test is used after scoring as a teaching instrument.

The Rearrangement Test.—This is an alternate response examination in which several items are grouped together as one exercise that is independent of other similar exercises in the test. The task is to number the items in each exercise in correct order. This test may be used either as a factual test or as a judgment, or reasoning, examination. In the former instance, the task is to arrange the items in the order of their occurrence, with historical dates or the steps in a particular routine; in the latter instance, to arrange them within each exercise in the order of their importance. Ordinarily in elementary grades, only factual types of items should be used in rearrangement tests; this type of examination, used in measuring judgment or reasoning, is quite difficult, and it requires ordinarily more information than is possessed by pupils in elementary grades.

Ordinarily, each exercise should include four or five items only. More than this number per exercise is uneconomical of time and tends to complicate scoring difficulties when pupils give incorrect orders for the items, whereas fewer than four items per exercise promotes success due to chance guessing. Only items that are readily identifiable should be used in this type of test, and unimportant facts and dates should not be used. Whatever facts are used should be distinct from others in the same exercise; for example, it is not defensible to include in one exercise the dates of two battles that occurred only a few weeks apart in the same year. It has been suggested that the items in each individual exercise always be presented alphabetically.

A sample partial rearrangement test follows.¹

¹ Devised by Agnes Harvill, Dudley School, El Paso, Tex.

SEMESTER EXAMINATION IN UNITED STATES HISTORY

Directions: Below are a number of exercises, each including five different facts. You are to number the first of the five in each exercise to occur 1; the second, 2; and so on. Look at the following example, which lists five battles fought in five wars in which the United States had a part:

3	The Battle of Antietam
5	The Battle of Château-Thierry
4	The Battle of Manila Bay
2	The Battle of New Orleans
1	The Battle of Yorktown

The numbers on the lines at the left of each item indicate the order in time in which the battles were fought. In each of the following exercises you are to arrange by number the five items in each exercise in the order in which they happened. Finish each exercise before you begin with the next one. Work rapidly.

-
1. _____ The Acquisition of the Florida Territory
 _____ The Acquisition of New Mexico and California
 _____ The Annexation of Texas
 _____ The Gadsden Purchase
 _____ The Louisiana Purchase

25. _____ The Attack on Fort Sumter
 _____ The Battle of Gettysburg
 _____ The Battle of Shiloh
 _____ The Surrender at Appomattox Courthouse
 _____ The Surrender of Atlanta

The rearrangement test is useful in subjects involving chronological dates, such as history, and in subjects involving processes, such as science and home economics (foods). This type of examination yields a definite indication of the precise facts at the command of the pupil in such situations as it is usable.

A limitation of the rearrangement test is that it is useful in only a few subjects of the curriculum. By far the most serious objection to it is its difficulty in scoring. Mechanically, the test is readily scored by means of a strip stencil, of course; the difficulty arises from the numerous displacements which occur in the correct order or arrangement of the items by number. These are in varying degrees of error. For example, if the correct order of five items is 1—2—3—4—5, and one pupil has his order 2—3—

4—5—1, he has four in correct order, each displaced by one number from its correct position, and one in incorrect order, displaced by four positions from the correct order. By various scoring formulas, the score of this exercise may be computed as any one of several values between -4 and $+4$. Obviously, the determination of a scoring formula is necessary, since the procedure in scoring and in computing scores will largely affect the final numerical score of each pupil. It is difficult to defend any formula used for this purpose.

The discussion heretofore has been concerned with the characteristics, advantages, and limitations of objective tests in general and with the presentation and discussion of the more common types of objective tests. Little has been said concerning the classroom teacher's procedure in preparing such examinations. Consideration must be given now to what the teacher does when planning for objective testing of academic achievement.

The Preparation of Objective Tests.—Peculiarly enough, the process of preparing objective tests begins even before instruction is given to the pupils who are to respond to the tests, and the work continues during the period of instruction. This may seem illogical, but it is entirely logical. The able classroom teacher collects subject-matter items that might be suitable for inclusion in an objective test while she is preparing her lesson plans for classroom teaching, and she continues assembling such material during the period of instruction. After a given period of instruction, the teacher should have selected many items of subject matter that might be used in an objective test. After several semesters of teaching the same subject, she should have these in addition to many other such items collected subsequently to those first selected, arranged in test-item form and placed in an indexed file, one item per card. The same subject-matter item might be used in several different types of tests, so it probably should have several cards, one for each type of test. The file for each type of test should be indexed according to subject and divisions within subjects. Each file card should show the wording of the item, the correct answer (for use in a scoring key), and the dates and number of all tests in which the item appeared. The teacher also should have a file of objective tests, with keys, and data sheets showing scores earned by pupils responding to each test. This type of preparation for testing, together with

constant revisions of the files, particularly in adding and subtracting items and in revising the wording of items as experience demonstrates their defects, should be done at all times. Careful preparation of this type makes it relatively easy for a classroom teacher to construct valid and reliable objective tests.

Whenever it becomes necessary for the teacher so equipped to plan and to prepare an objective test, she must first decide on the scope to be covered by the examination. This means that a decision must be made as to how much subject matter will be covered by the test and how long the test itself will be, what types of objective tests will be used as divisions of the completed examination, and how many items must be included in each subdivision of the examination. It is difficult to determine exactly how much time must be given to each division of the total test and so to the total test. The rate of work will vary on different types of objective tests with different pupils and also on identical examinations with different pupils. Rate, in general, varies also with mental ability and with differences in the difficulty of the test content. The test maker must learn from experience to gauge the working time needed on objective tests. There is, of course, the option of "trying out" a test, stopping the pupils' work when 90 per cent have finished their work. Thereafter the given test may be timed rather accurately, as well as improved by eliminating any item or items that for any reason are unsatisfactory. There is a possibility that tests so prepared and used may become static; care must be exercised constantly to revise the measuring instruments and to maintain their efficiency in evaluating academic achievement.

The test maker must remember to use more than one type of objective test in each examination; some pupils respond more successfully to one type of question situation than to another. In lower grades, tests must be short, possibly including only 25 to 30 items in all. The length of tests may be increased to 80 to 100 items in upper elementary grades.

Since the selection and editing of test items will have been accomplished when test items are placed in the test files, it is necessary for the test maker only to isolate the cards for the items to be used in a given examination. From one-fourth to one-third more items should be chosen than will be used in the final form of

the test. These should be arranged by the examiner roughly in their order of difficulty, and the final selection made from the total group, eliminating any that for any reason are not suitable. Following the identification of the specific test items, there remains the preparation of the instructions for the test and of the scoring keys. The instructions to the examiner should cover the examiner's procedure both in giving the test and in scoring the completed examinations.

Where economy of teacher time or of money is a factor, it is possible to make error analyses covering pupil responses to any given test and to determine the suitability of each item in terms of the percentages of error of all items. In addition, the adequacy of the scoring key should be checked. When all unsuitable or inaccurate materials have been eliminated, other items may be added to return the test to its desired length, and preferably these items should be from the files and known to be adequate in every way. The test may then be prepared in mimeographed form for use in several successive semesters. Standard score sheets may be prepared for the pupil responses, in order that the tests themselves may be retained for use again and again. This makes some difficulty in using the tests for remedial teaching purposes, and it requires extreme care to protect the content of the test from the knowledge of pupils who may take it later. These things can be done, however, when it is necessary to do them to achieve economy in measurement.

All scoring keys should be carefully checked for accuracy during tryout, and all alternate correct answers should be added to the key. The use of standard scoring sheets makes it possible to prepare stencils for each test individually. These stencils and the scoring key should be filed with the test itself. The effort should be to use only window and strip stencils, as these are the most effective in securing economy in scoring time.

Finally, always whole credit, one point for each correct response, should be given. As noted in the previous discussion of individual types of tests, scoring formulas must be utilized, *e.g.*, in determining the final score on a true-false test. These, however, are based on whole credit points for all correct responses. No attempt should be made to give more credit on more difficult items than is given on easier items.

Concluding Statement.—The materials in this chapter have dealt in general with objective tests and specifically with certain commonly used types of objective tests. The teacher's procedures in planning for and accomplishing the preparation of objective tests have been discussed. It should not be assumed that this discussion covers completely all that might be said concerning this subject; this is not true. Entire volumes have been written concerning the subject. What has been done in this chapter is to present a brief discussion covering the important aspects of the preparation and use of objective measuring instruments. The student will find it necessary to read widely on this subject in the references for further study.

Readings for Further Study

NOTE.—These references include only textbooks and bulletins and monographs. Provision is made in the exercises to insure the reading of periodical materials. So many magazine articles have been published that it would be difficult to present a fair sampling of the whole. The most important of the following references have been indicated by an asterisk.

1. B. R. BUCKINGHAM, *Research for Teachers*. Chap. VI. New York: Silver, Burdett & Company, 1926.
2. N. FENTON and D. A. WORCESTER, *Educational Measurements*. Chap. IV. Boston: Ginn and Company, 1928.
3. C. W. ODELL, *Traditional Examinations and New Type Tests*. Chaps. III, IV, VII, IX–XVI. New York: D. Appleton-Century Company, Inc., 1928.
4. J. S. ORLEANS and G. A. SEALY, *Objective Tests*. Chaps. III–XIII. Yonkers-on-Hudson, N. Y.: World Book Company, 1928. For scanning only.
- *5. H. D. RINSLAND, *Constructing Tests and Grading in Elementary and High School Subjects*. Consult the index.
- *6. G. M. RUCH, *The Objective or New-type Examination*. Consult the index.
- *7. C. RUSSELL, *Classroom Tests*. Chaps. II–VII. Boston: Ginn and Company, 1926.
8. C. RUSSELL, *Standard Tests*, Chaps. IV, V, VI.
9. E. W. TIEGS, *Tests and Measurements for Teachers*. Chaps. IV, V.
10. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. XIX.
11. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chaps. XVIII, XXIV.

Printed Materials Other than Books.

1. J. F. BURSCH and H. MELTZER, *The New Examination—Its Construction and Use*, *Bulletin 422*. Corvallis: Oregon State College, 1926.

2. G. F. MILLER, *Objective Tests in High School Subjects*. Norman, Okla.: The Author, 1926.
3. J. S. ORLEANS, Manual on the Local Construction and Uses of Objective Tests, *Bulletin* 893. Albany: University of the State of New York, 1927.
4. D. G. PATERSON, *Preparation and Use of New-type Examinations*. Yonkers-on-Hudson, N. Y.: World Book Company, 1925.

Questions and Exercises

1. What is the nature of the evidence concerning the comparative reliabilities of essay and new-type examinations? Is the evidence conclusive? What are the advantages of each of these types of tests?

2. In what types of objective tests are there subjective factors that affect the final scores earned by the pupils responding to the tests? Prepare a list of such subjective factors. State your reasons for including each item in the list.

3. Study the various ways of presenting true-false items in an objective examination. What are the advantages and disadvantages of each of these methods of presentation? Consider similarly multiple-response items, completion items, and matching exercises.

4. What are the factors that affect the validity of a teacher-made, objective test? In just what way does each of these factors affect validity?

5. Is it possible to say positively just how many items of a specified type should be answered in a given time interval by pupils in any one school grade? Why?

6. Are all types of objective test exercises equally useful in all school grades? State your reasons for your reply, and indicate which types of test items are probably most useful in several different school grades (if useful at all in the selected grade levels).

7. What practical objections (cost and time requirements in preparation and the like) can be brought against any of the different types of objective tests described in the preceding chapter discussion?

8. Make a study of teacher-made objective tests that have been devised for use in elementary-school classrooms. It is desirable that the instructor have a file of such measuring instruments for student use. The student should prepare a two-column (paired) analysis showing the excellences and the defects of each test examined. In making this analysis, reference should be made to specimen tests given in

H. D. RINSLAND, *Constructing Tests and Grading in Elementary and High School Subjects*.

G. M. RUCH, *The Objective or New-type Examination*. Chap. IX.

G. M. RUCH and G. D. RICE, *Specimen Objective Examinations*.

Limitations of space prevent giving many complete objective examinations, but two of varying degrees of excellence are given completely. These may be used if others are not available for the doing of this exercise.

TEST A. GEOGRAPHY OF THE BRITISH ISLES

Grade 5

Directions: Fill the blanks in the sentences.

1. The British Isles are _____ of the continent of Europe.
2. The Island of Great Britain includes _____, _____, and _____.
3. All British possessions in all parts of the world are called _____.
4. The wheatlands of England are in the _____ part.
5. Sheep are raised in the _____ part of England.
6. _____ and _____ are raised in the highlands of Scotland.
7. Important manufactures of England are _____, _____, _____, and _____.
8. Four imports of the British Isles are _____, _____, _____, and _____.
9. Four exports of the British Isles are _____, _____, _____, and _____.
10. _____ is the largest city in the world.
11. The two most important seaports in England are _____ and _____.

Directions: Match the words in the two columns.

- | | |
|--------------------------------------|-----------------|
| 1. A shipbuilding center | Edinburgh |
| 2. An iron and steel center | Pennines |
| 3. A river in Scotland | Sheffield |
| 4. Seaport on the English Channel | Leeds |
| 5. Seaport in Wales | Manchester |
| 6. Island in the British Channel | Dogger Bank |
| 7. Mountains | Stoke-on-Trent |
| 8. A city important for printing | Southampton |
| 9. A cotton-manufacturing center | Cardiff |
| 10. A woolen-manufacturing center | Clyde |
| 11. An old prison | Birmingham |
| 12. A mountain peak | Jersey |
| 13. A great pottery center | Tower of London |
| 14. Famous fishing grounds | Glasgow |
| 15. Center for the making of cutlery | Ben Nevis |

TEST B. COLONIZATION OF NORTH AMERICA

Name _____ Date _____

Part I

Directions: Write the correct answer in the blank at the right.

1. What was the first European settlement in what is now the United States?.....
2. Who was the greatest of the "sea dogs"?.....

3. What was the first permanent English settlement in America?..... _____
4. What were the names of two English companies formed to colonize America?..... _____
5. What three important things happened in 1619?..... _____
6. Who was the first governor of the Plymouth Colony?... _____
7. What were the Plymouth settlers called?..... _____
8. What people founded New York?..... _____
9. Which colony did the Swedes settle?..... _____
10. What is the oldest University in the United States?.... _____
11. Who was the British general in the French and Indian War?..... _____
12. Who was his aide?..... _____
13. What treaty closed the French and Indian War?..... _____
14. Who was the "Father of New France"?..... _____
15. Who was the ruler of England during the greatest period of English history?..... _____
16. Who was Queen Elizabeth's favorite courtier?..... _____
17. What were two products that were taken to England from the colonies?..... _____

18. What were three kinds of labor used in the colonies?.... _____

19. What did England defeat, thereby becoming "Mistress of the Seas"?..... _____

Part II

Directions: In the parentheses, write the letter of the answer that makes the statement correct.

- () 1. The French came to the New World.
 - a. to Christianize the Indians.
 - b. to trap animals for their fur.
 - c. to trade.
 - d. to find new homes.
- () 2. The first English Colony was established in
 - a. 1619.
 - b. 1620.
 - c. 1607.
 - d. 1600.
- () 3. Maryland was
 - a. a royal colony.
 - b. a proprietary colony.
 - c. formed by a stock company.
 - d. sponsored by businessmen.
- () 4. The New England Confederacy was formed
 - a. for protection against the Indians.
 - b. for trading purposes.

- c.* to further religious worship.
 - d.* to celebrate the passing of the first winter.
- () 5. Georgia was founded by
 - a.* Oglethorpe.
 - b.* Bradford.
 - c.* The Earl of Shaftesbury.
 - d.* John Smith.
- () 6. The Acadians were
 - a.* French people living in Nova Scotia.
 - b.* a religious group.
 - c.* people living on an island near Florida.
 - d.* governors of the English colonies.
- () 7. The leader of the Jamestown colony was
 - a.* Pocahontas.
 - b.* John Smith.
 - c.* Sir Gilbert.
 - d.* Walter Raleigh.
- () 8. The most successful colonists were the
 - a.* English.
 - b.* Swedes.
 - c.* French.
 - d.* Spanish.
- () 9. Massachusetts Bay Colony was established
 - a.* to gain land.
 - b.* to gain riches.
 - c.* for religious purposes.
 - d.* to hunt.
- () 10. Henry Hudson explored for
 - a.* England.
 - b.* Holland.
 - c.* Spain.
 - d.* France.
- () 11. William Penn was a
 - a.* Pilgrim.
 - b.* Puritan.
 - c.* Quaker.
 - d.* Baptist.
- () 12. The greatest French explorer was
 - a.* Cartier.
 - b.* Champlain.
 - c.* La Salle.
 - d.* Montcalm.
- () 13. The treaty that closed the French and Indian War was signed in
 - a.* 1758.
 - b.* 1763.
 - c.* 1783.
 - d.* 1800.

- () 14. After the French and Indian War, the country that gave up all claim to America was
- France.
 - England.
 - Spain.
 - Sweden.

Part III

Directions: Match the items in the left-hand column with those in the right-hand column. There are more in the left-hand column than in the right-hand column, so do not expect to use all of the items in the left-hand column.

- | | | |
|-----------------------|-----|---------------------------------------|
| 1. Governor Bradford | () | Queen Elizabeth's favorite courtier |
| 2. Peter Stuyvesant | () | Founder of the Carolinas |
| 3. Sir Francis Drake | () | Governor of New York |
| 4. Sir Walter Raleigh | () | Founder of Rhode Island |
| 5. Lord Ashley Cooper | () | Governor of Plymouth |
| 6. James Oglethorpe | () | Leader of the French at Quebec |
| 7. Roger Williams | () | Most famous of the English "sea dogs" |
| 8. Montcalm | () | First proprietary colony |
| 9. Maryland | () | Leader of the English at Quebec |
| 10. Wolfe | () | Founder of Georgia |
| 11. Braddock | | |
| 12. Pennsylvania | | |

Section 2, Part III

- | | | |
|----------|-----|--------------------------------------|
| 1. 1607 | () | Landing of the Pilgrims |
| 2. 1763 | () | First legislature met |
| 3. 1608 | () | Founding of Massachusetts Bay Colony |
| 4. 1619 | () | Founding of Jamestown |
| 5. 1620 | () | Charter for the Carolinas |
| 6. 1732 | () | Treaty of Paris |
| 7. 1630 | () | Founding of Georgia |
| 8. 1662 | () | Founding of Quebec |
| 9. 1665 | | |
| 10. 1783 | | |

Part IV

Directions: Write the answer in the blank that you think best answers the question.

- The colony formed for the home for debtors was _____.
 - A colony established by a king is called a _____ colony.
 - A colony established and governed by an individual is called a _____ colony.
 - The oldest University in the United States is _____.
 - The Father of New France is _____.
 - By the treaty that closed the French and Indian War _____ gave up all land east of the Mississippi River and Canada to _____.
- The country of _____ does not own land in North America now.

7. The _____ people made the best colonists because _____

8. The French made poor colonists because _____

9. _____ is the oldest European settlement in the United States.

10. Quebec was founded by _____.

11. Two colonizing companies were the _____ Company and the _____ Company.

12. Why were the following founded:

Jamestown _____

Massachusetts Bay Colony _____

Pennsylvania _____

New York _____

Plymouth _____

Rhode Island _____

9. Prepare a series of reports of periodical articles dealing with the construction or the use of objective tests. The following sources may be consulted in obtaining titles of suitable articles:

a. Periodical indexes, such as the *Reader's Guide to Periodical Literature* and the *International Index to Periodicals* for older articles and the *Education Index* for recent articles.

b. *Education Abstracts* and *Psychological Abstracts*.

c. *The Loyola Educational Digest*.

d. Bibliographies at the ends of chapters in standard textbooks dealing with educational measurements.

A standard form should be used in preparing the reports. One good form is indicated below:

Page 1. The heading (one-third page) and space for tabular data (two-thirds page). Only significant tabular data should be reported.

The heading follows:

FORM FOR PERIODICAL REPORTS. Report No. _____ Date _____ Score _____

Your name _____ Course _____

Department _____ Number _____ Section _____

Reference (Title of article, etc. Use standard form)

Page 2. Graphs and charts (one-half page) and a summary, or *précis*, indicating briefly all significant information in the article. The summary should be as complete as reports in *The Loyola Educational Digest*, etc.

Page 3. The author's conclusions (verbatim; one-third page) and the student's own specific comments and criticisms, including any quotations, etc., to support the comments and criticisms (two-thirds page).

EXAMINATION ON THE CONSTRUCTION AND USE OF OBJECTIVE TESTS

Directions: The following statements are to be completed with one or more words. Supply the missing word or words that will best complete the

statement in the blanks. If you do not know an answer, skip the item and continue with the next one. Work as rapidly and as accurately as you can. DO NOT GUESS!

1. The type of test item (technique or mechanical form) should be decided principally upon the _____ of that technique to the particular item of subject matter.

2. In writing instructions for a test, phrase the directions to meet the level of the _____ mentalities in the group.

3. In scoring objective tests, it is best to score items only for _____ and not for _____.

4. The perfect matching examination is bad because it is likely to measure only _____.

5. The most expensive objective test to prepare and use is the _____.

6. The type of objective test that makes use of pictures is the _____ examination.

7. One advantage of the simple recall or completion test is that it allows some freedom of _____.

8. Scoring is said to be highly _____ when the answer is either right or wrong and when the scorer's opinion does not enter into the marking.

9. The test that is most nearly perfectly objective is the _____ examination.

10. If the items in a test are arranged in order from easiest to hardest, the _____ and _____ of the test usually are increased.

11. Variations that occur when equally competent persons mark the same examination are called _____.

12. By two sets of measures of the same ability or function is meant _____.

13. The "best examination" for the present must be that one which by experiment can be shown to be most _____.

14. _____ should sample the kinds of activities that occur daily in the classroom.

15. _____ is a prime essential for reliability of measurement.

16. The most difficult type of examination question to build is the _____.

17. The difference between a thought and a memory question does not reside in the question itself but in the _____ of the pupil.

18. Intensive sampling is represented by the _____ type of test.

19. _____ items compromise between the completion and recognition types of items.

20. The fact that some statements are bound to be ambiguous or controversial are limitations especially disadvantageous to the _____ item.

21. _____ test items should follow rules of good language expression, grammar, spelling, punctuation, and capitalization.

22. The principal objection to the essay examination reduces to the question of _____ of scoring.

23. The two main characteristics of reliability are _____ and _____.
24. When two sets of measures of the same ability or function are correlated, we term the resulting coefficient of correlation a _____ coefficient.
25. In the _____ test, a number of statements are presented with some of the main items omitted, which are to be supplied by the person responding to the test.
26. A _____ test is one that yields detailed information concerning pupil achievement in one or more relatively restricted fields.
27. A test score that may be influenced by the personal opinion of the scorer is said to be _____.
28. Standardized tests should not be the complete measurement program because they are too _____ and therefore do not fit _____ conditions.
29. _____ blanks are more convenient for scoring than are _____ blanks.
30. A test may be said to be _____ when it measures adequately what has been taught.
31. The two main divisions of objective examinations are _____ and _____ tests.
32. In the _____ test, there are two columns of facts, and the pupils responding are asked to relate those in one column to those in the other.
33. The essay or _____ test gives experience in language-usage training.
34. In a true-false test, _____ of sampling is possible in _____ time.
35. In three response tests, the formula for scoring is _____.
36. Matching tests may be used either as _____ or as _____ examinations.
37. The best answer, or judgment, tests may be regarded as _____ forms of _____ or _____ tests.
38. The test that offers several alternate answers from which to choose is called the _____ test.
39. Objective examinations measure _____ more than _____.

Part II

Directions: Each of the following statements is either true or false. If the statement is true, place a check on the line before the letter T; if false, before the letter F. If you do not know which is the correct answer, skip the item, and continue with the next one. DO NOT GUESS! Work as rapidly and as accurately as you can.

- _____ T _____ F 1. Tests are best validated by the judgment of competent persons.
- _____ T _____ F 2. Checking, underlining, encircling etc., are more economical of scoring time than are words written in blanks.

- _____ T _____ F 3. Supposed equally difficult tests are not always equally difficult to those taking them.
- _____ T _____ F 4. A test may be reliable and yet not be valid.
- _____ T _____ F 5. Validity refers to the degree of consistency of measurement.
- _____ T _____ F 6. If three techniques of measurement are used in one test, it is necessary to have three sets of instructions.
- _____ T _____ F 7. Subjectivity and limited sampling are closely related.
- _____ T _____ F 8. In general, the larger the number of equivalent forms the greater is the usefulness of the test.
- _____ T _____ F 9. If a test is totally unreliable, it has low validity.
- _____ T _____ F 10. The validity of a test is increased by having the easiest items first and the hardest ones last.
- _____ T _____ F 11. In completion questions, the answers should be placed in blanks that always occur at the right end of the line.
- _____ T _____ F 12. It is occasionally necessary to sacrifice grammatical construction in the interest of clarity in test instructions.
- _____ T _____ F 13. The instructions should direct the pupil where and how to record his answers.
- _____ T _____ F 14. "Culling out" should be done the same day that the items are written, so that the ideas may be clearly in mind.
- _____ T _____ F 15. Items should be chosen because they cover the subject thoroughly but at the same time so as to avoid the trivial points.
- _____ T _____ F 16. The mental set in the ordinary political, economic, or social argument approaches most nearly the true-false attitude.
- _____ T _____ F 17. A norm is a value used in interpreting or transposing a score.
- _____ T _____ F 18. The majority of the standard tests are sufficiently valid and reliable.
- _____ T _____ F 19. Differences in subject matter, and relative economics and reliabilities of different types of test items, may affect the reliability of a standardized test.
- _____ T _____ F 20. In making test instructions, use the simplest synonyms for all words or ideas expressed.
- _____ T _____ F 21. The examination should be a measure of ability to understand and follow directions as well as of achievement.
- _____ T _____ F 22. It is possible for the validity of a test to be somewhat greater than its reliability.
- _____ T _____ F 23. Tests should measure the understanding of the question as well as achievement.

- _____ T _____ F 24. In writing test instructions, they should be phrased so as to be on the level of the average student.
- _____ T _____ F 25. "Perfect matching" is superior to "imperfect matching."
- _____ T _____ F 26. In constructing completion items, it is best to copy a passage directly from the text, omitting a few critical words.
- _____ T _____ F 27. Items that permit only slight differences due to scorers' prejudices are called objective.
- _____ T _____ F 28. A thought question is good because it may be used later to retest the same pupil and show reasoning advancement.
- _____ T _____ F 29. It is well to use double negatives in constructing true-false items, as they require careful and clear thinking.
- _____ T _____ F 30. Instructions will tell whether to work rapidly or to take your time; to guess or to leave blanks.
- _____ T _____ F 31. In a true-false examination, the best way to mark the answer is to have a dotted line to the left of the number on which the student marks T or F or plus or minus.
- _____ T _____ F 32. The greater the concentration of scores lying adjacent to the score distribution mean the more valid the test.
- _____ T _____ F 33. A test is said to be valid but unreliable if it discriminates between different levels of ability.
- _____ T _____ F 34. Terminal blanks (as in the simple recall test) should be placed immediately after the close of the statement.
- _____ T _____ F 35. Only one type of question should be used in any objective examination.
- _____ T _____ F 36. If the items in a test are arranged in an increasing order of difficulty, the reliability of the test will be increased.
- _____ T _____ F 37. The important factors in test instructions are clarity, brevity, and accuracy of statement.
- _____ T _____ F 38. It has been found that a test score is more valid when pupils are instructed not to guess, when the answering of the items permits guessing with equal chances of success and failure.
- _____ T _____ F 39. It is wise to study standardized tests before making objective examinations.
- _____ T _____ F 40. The response to a simple recall-test item should be very short.
- _____ T _____ F 41. Multiple-response examinations often measure reasoning when the response items are long statements.

- _____ T _____ F 42. A rearrangement test is usually less reliable than is a true-false test.
- _____ T _____ F 43. Matching tests may be used to measure either factual mastery or judgment.
- _____ T _____ F 44. Before devising an examination, the test maker should know the capacities or abilities to be measured by the test.
- _____ T _____ F 45. An advantage of the best answer, or judgment, test is that it allows the testing of ability to apply principles to concrete situations.
- _____ T _____ F 46. The identification test is not completely objective as to scoring.
- _____ T _____ F 47. A diagnostic test indicates general limitations of pupils in a given subject field.
- _____ T _____ F 48. One of the advantages of the rearrangement test is that it is easily and rapidly scored.
- _____ T _____ F 49. The recall, or completion, types of test are easier to construct than is the true-false test.
- _____ T _____ F 50. One method of dividing the items into two groups for use in equivalent forms of an objective test is by chance.

CHAPTER VII

STANDARDIZED TESTS IN THE LANGUAGE ARTS

Man's social heritage is transmitted from generation to generation and communicated from one individual to another in any given generation by means of words, pictures, and symbols. Words, in particular, are employed as the medium for communicating ideas. If they are to serve this purpose, they must be understood both by the individual endeavoring to communicate the idea and by the person to whom the idea is being communicated. Several different uses of language are involved in the communication of ideas. One conveys his ideas to others by speaking or by writing and receives the ideas of others by hearing them speak or by reading what they have written.

Modern language is elaborate. Each individual must master the tools and modes of expression and of communication, and this cannot be done by imitation alone. The school gives a large part of its time to developing the vocabulary of its pupils and their mediums of expression, attempting to develop their handwriting, spelling, grammar and composition, and reading. It attempts to introduce and to induce mastery of the literature of the past and of the present as well as words and forms of expression not encountered ordinarily in the home or in the daily environment of pupils outside the classroom.

The purpose of this chapter is to present information concerning the standard tests that have been developed for use in measuring pupil achievements in the language arts: in reading, handwriting, spelling, grammar, mechanics of expression, and composition. The language arts are important tool subjects; and because of this importance, means for measuring achievement must be available in all school grade levels. The information yielded by such measuring instruments provides for the proper guidance of pupil learning in mastering this indispensable vehicle of expression and communication. It should be noted, perhaps, that only standardized tests will be discussed in this chapter but

that teacher-made objective tests can be devised and used for all purposes for which standardized tests are valuable. The only advantage of the standardized tests lies in their margin of refinement over carelessly constructed teacher-made objective tests and in the added information that a teacher may glean from their norms. In general, the same comment might be applied also to the discussions of the following three chapters.

The organization of this chapter presents in order handwriting, spelling, composition, mechanics and grammar, and reading.

I. HANDWRITING

The purpose of handwriting is to develop in the pupil skill in the use of conventional symbols so that he can record facts and ideas for communication or for permanent keeping. Since handwriting is the medium for preserving the achievements of the mind, it serves also to compel an individual to clarify and to organize his thinking. One is often unaware of a lack of information or of a confusion in his ideas until an attempt is made to write a meaningful statement. Achieving an awareness of this situation often serves as an incentive for obtaining accurate, definite, and complete information. Handwriting, then, serves a dual purpose: It motivates individuals in the acquisition and organization of ideas and facts, and it serves as the means for recording them. Efficiency in activities in which handwriting is used requires both legibility and speed of performance; and handwriting is of use both in school and in work outside the school. Mechanical devices, such as the typewriter, have not displaced handwriting as an essential tool for achievement in the repertoire of the average individual.

Skills Necessary in Handwriting.—Two independent aspects must be found in successful achievement in handwriting: The individual must write legibly and speedily. There is no positive relationship between legibility and speed of handwriting.¹ Both must be measured in order to evaluate handwriting achievement properly. In so far as legibility or handwriting quality is concerned, six factors are recognized: letter formation, alignment, spacing, slant, line quality, and size. The relative importance of these factors is indicated by the values given on page 148, the

¹H. B. REED, *Psychology of Elementary School Subjects*, pp. 275-278, Boston: Ginn and Company, 1938.

average judgments of 47 experienced teachers and supervisors of handwriting in answer to the question: "If you should have 100 points to distribute among the six elements of handwriting—alignment, letter formation, line quality, spacing, size, and slant—how many points should you designate for each element?" The following were the results:

Letter formation.....	34.67
Alignment.....	13.70
Spacing.....	13.68
Slant.....	13.23
Line quality.....	12.87
Size.....	11.85
Total.....	100.00*

* These data were furnished by Irene Prewit, W. S. Benson and Company, Austin, Tex.

Measuring Speed of Handwriting.—An excellent example of how speed in handwriting is measured occurs in the Gettysburg Edition of the Ayres Measuring Scale for Handwriting.¹ This is a revised edition of a scale first published in 1912. Use is made of the first three sentences of Lincoln's Gettysburg Address. The pupils read and copy this material until they are familiar with it. They then copy it in ink on ruled paper, beginning at a given signal and writing for precisely two minutes. The copy with the count of the letters is as follows:

Four 4 score 9 and 12 seven 17 years 22 ago 25 our 28 fathers 35 brought 42 forth 47 upon 51 this 55 continent 64 a 65 new 68 nation 74 conceived 83 in 85 liberty 92 and 95 dedicated 104 to 106 the 109 proposition 120 that 124 all 127 men 130 are 133 created 140 equal 145. Now 148 we 150 are 153 engaged 160 in 162 a 163 great 168 civil 173 war 176 testing 183 whether 190 that 194 nation 200 or 202 any 205 nation 211 so 213 conceived 222 and 225 so 227 dedicated 236 can 239 long 243 endure 249. We 251 are 254 met 257 on 259 a 260 great 265 battlefield 276 of 278 that 282 war 285.

Speed is determined in terms of the number of letters formed per minute.

Measuring Quality of Handwriting: Survey Scales.—Survey scales are available for use in evaluating both manuscript, or print, writing and script, or cursive, writing. In many elemen-

¹ Devised by L. P. Ayres and published by the Public School Publishing Company, Bloomington, Ill.

tary schools, manuscript writing is taught in primary grades, since the performance requirements for manuscript writing are fewer than are those for cursive writing. For example, coordination and control of muscular activity is not so important in manuscript as in cursive writing. Ordinarily, the transition from manuscript to cursive writing is made in grade 3.

The first handwriting scale considered was devised for use in evaluating achievement in manuscript writing; the following survey scales, for use in evaluating cursive writing.

The Conard Manuscript Writing Standards.—There are two scales for measuring the quality of manuscript writing: One is used in scoring pencil writing; the other, in scoring pen writing.

Dear Miss Conard
I am glad that we can
our writing.

. The . . . Peddler

. I wish I lived in a caravan,

. With a horse to drive,

. . . like a peddler-man!

. Where he comes from nobody

: . . knows

. Or where he goes to,

. . . . but on he goes. . . .

The specimens included in both scales were chosen from 5,000 samples of manuscript writing. The author says of both scales:

They are spread out in steps showing progress in development of form, size, spacing of letters, words, and lines, quality of line, and

arrangement on the page as a whole. Other factors such as muscular control, facility of movement, and proper use of tool were also considered in selecting these samples.¹

The first sample on page 149 is from the chart for *Pen Forms*; the second, from the chart for *Pencil Forms*. The scale for *Pen Forms* includes 10 specimens; that for *Pencil Forms*, 12 specimens. The pencil specimens range from undecipherable writing of a first-grade child to good writing of a fourth-grade child. The pen specimens range from beginning inkwork of a third-grade child to good inkwork of a sixth-grade child, with two older hands to bring up the range. The specimens used for illustration are chosen near the middle of the two scales.

The author of the scale offers the following information and directions concerning the judging of manuscript writing, using the scale:

Numerically the standards for both pencil and pen forms are based on a 100-point scale, consisting of 20 steps of 5 points each. The rating of pencil forms extends to point 60 on the scale, and the rating of pen forms begins at 50 and extends to 100. The pencil and pen forms therefore overlap between 50 and 60.

Pen		Pen	
Sample	Score	Sample	Score
1	50+ to 55	6	75+ to 80
2	55+ to 60	7	80+ to 85
3	60+ to 65	8	85+ to 90
4	65+ to 70	9	90+ to 95
5	70+ to 75	10	95+ to 100

To score any paper, find among the standards above the sample to which it most closely corresponds both in quality and in description. If there are slight variations, score within the 5 points. In most cases it will simplify the scoring if the mid-point of the interval is used. For example, for sample 2 (pen) use the score 57, and so on.

In judging any handwriting, the factor of first importance is legibility or readableness; that is, the letter forms should be simple, with no

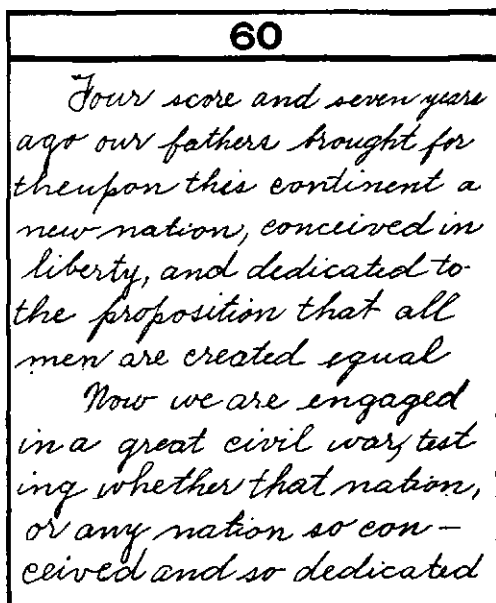
¹ EDITH U. CONARD, *Conard Manuscript Writing Standards*, New York: Bureau of Publications, Teachers College, Columbia University.

unnecessary parts, and they should be distinct, each letter form having its chief characteristic definitely noted. The letters should be compact so that the words stand out as units.

The factor of second importance is the beauty of the handwriting. The form of the letters should be as distinct as possible, considering age and experience of the writer. There should be good proportion in the letter forms, with no exaggerated strokes. The letters should be uniform and regular in both shape and slant. The writing should be so arranged on the page as to secure good spacing of letters, of words, and of lines.

The factor of third importance is character, which is influenced by the correct use of suitable tools. These, when rightly handled, give characteristic results. Facility and freedom in handling tools result in definite strokes and boldness of form. Individuality appears as a result of characteristic strokes and form.

The Ayres Measuring Scale for Handwriting, Gettysburg Edition.
This scale is based upon legibility as shown by the time required



to read the samples. It includes eight specimens of handwriting, uniform in material, since the first three sentences of Lincoln's Gettysburg Address are used as copy. The specimens are scaled uniformly, with values ranging from 20 to 90. The method of

obtaining the samples of pupils' handwriting has been indicated on page 148, in connection with the discussion of the measuring of handwriting speed.

A specimen section of the scale of approximately median quality is shown on page 151.

The author's directions as to the scoring of pupils' samples of handwriting for quality follow:

To score samples slide each specimen along the scale until a writing of the same quality is found. The number at the top of the scale above this shows the value of the writing being measured. Disregard differences in style, but try to find on the scale the quality corresponding with that of the sample being scored. With practice the scorer will develop the ability to recognize qualities more rapidly and with increasing accuracy. If the scoring is done twice, the results will be considerably more accurate than if done only once. The procedure may be as follows: Score samples and distribute them in piles with all the 20's in one pile, all the 30's in another, and so on. Mark these values on the backs of the papers, then shuffle the samples and score them a second time. Finally make careful decisions to overcome any disagreements in the two scorings.

48

I will play with the two balls.

Is this the north wind.

Do not throw the red ball.

Every work has work to do.

Norms are available for use in grades 2 through 8 for both speed and quality of handwriting. All directions and pertinent data appear on the scale itself.

*The Leamer Scale for Grade Standards in Quality.*¹—This scale was devised for use with *Leamer's Practice Sentences in Handwriting*.

¹ Devised by E. W. Leamer and published by the Public School Publishing Company, Bloomington, Ill.

ing. The scale items include in each four sentences for each quality. The quality levels occur by fives between 30 and 70 inclusive. A sample of the scale materials of median quality value is shown on page 152.

The time limit for writing the pupils' samples of handwriting is uniform for all grades, 2 minutes. Different sentences may be used as copy. The standards given by the author of the scale for both speed and quality follow:

Grade	Quality	Rate, times
2	35	3
3	40	4
4	45	5
5	50	6
6	55	7
7	60	8
8	65	9

It should be noted that the scale materials are grouped on separate small scale sheets, with three samples only appearing per grade. Rate is determined roughly by the number of times the pupil writes the given sentence in the two minutes allowed for obtaining the samples of penmanship. Obviously, this scale does not give such accurate information as does the Ayres Measuring Scale for Handwriting, but it is more closely related to classroom instruction materials and procedures than is the Ayres scale.

The student of educational measurements should not fall into the error of thinking that the two cursive survey scales that have been presented represent all the productive work that has been done. They are typical of the many scales available. All the available scales are used in about the same manner and for the same general purposes. The data in Table V (page 69) indicate that different handwriting survey scales validate each other satisfactorily. The conclusion is that regardless of which scale is used, similar results concerning the quality of pupils' handwriting will be obtained.

Survey measuring scales for handwriting are essentially scoring instruments; they are not tests. They are used to establish uniform quality values for given specimens of pupils' handwriting, *i.e.*, more valid and reliable judgments than could be obtained

without some standardized "measuring stick." The use of such instruments lends relatively little information about the handwriting of any individual pupil, since it indicates merely a level of quality, not what factors have contributed to that quality or what defects in handwriting performance have restricted the quality of the specimen. In general, results obtained by using such a survey scale permit comparisons between groups of pupils as to handwriting achievement or comparisons for a given group at different times, although it is possible to make roughly accurate comparisons for a given pupil from time to time. The major usefulness of the findings of a survey scale is to indicate the existence of a general level of handwriting achievement in a given class or classes and to indicate those pupils whose handwriting performance is below acceptable standards. When such a scale is used in this way, it provides information necessary to the use of diagnostic handwriting measuring materials. Remedial teaching, in particular, should be based on diagnostic testing.

Diagnostic Materials for Measuring Quality of Handwriting.—The diagnosis of handwriting defects and excellences involves the investigation of the product of handwriting rather than the observation and analysis of handwriting performance. It is possible that eventually motion picture records of handwriting in process may be used as a performance test; but at present, diagnosis is accomplished by utilizing samples of pupil penmanship. Several devices have been made available, including both scales and analysis charts. These are discussed in the following paragraphs.

*The Freeman Chart for Diagnosing Faults in Handwriting.*¹—This is one of the earlier and more useful instruments of its type. Included in it are three qualities of performance—poor, average, and good—for five characteristics of handwriting—slant, alignment, quality of line, letter formation, and spacing. The poorer quality is given a score of 1; the average, a score of 3; and the better quality, a score of 5. If the sample of handwriting being scored seems to lie between any two of these values, it can be given an intermediate score value of 2 or 4.

The purpose of this chart is to focus the scorer's attention on the excellences and the defects in handwriting to be found in the

¹ Devised by F. N. Freeman and published by Houghton Mifflin Company, Boston.

sample under consideration rather than on the quality of the handwriting. Obviously, this scale serves to accomplish this purpose more than does a survey scale, but more detailed analysis is required than ordinarily results before specific instruction can be planned for individuals. For example, it is not enough to know that a given pupil has a letter-formation score of 1; it is essential to check each letter specifically to ascertain the defects, if any, in the formation of specific letters before any adequate remedial work can be undertaken.

*The Pressey Chart for Diagnosis of Illegibilities in Handwriting.*¹ This chart presents a scheme for diagnosing the specific elements that interfere with legibility in the handwriting of an individual. Its authors say that the major purpose of handwriting is not primarily to present an attractive appearance but to be read and that efforts for improvement should be directed specifically toward the elimination of any formations that interfere with legibility.

The chart is based on an analysis by 16 readers of 800 samples of handwriting, obtained from pupils in grade 3 through the college level. Each reader was given papers from all school levels and told to read them, checking any places where he had to stop (even momentarily) to ascertain the meaning. The formations that caused difficulty were then ascertained and classified. A total of 329 such different forms were noted; in all, 2,736 illegibilities were dealt with during the study. Many of these fall into rather general groupings.

The list that the authors incorporated in their chart includes such items as the following:

Words crowded
 Lines crowded
 Too angular
 a like o
 d like cl
 e too high
 s indistinct
 v like n

The authors suggest that a piece of handwriting produced in the course of regular classwork, such as a theme or examination,

¹ Devised by S. L. PRESSEY and L. C. PRESSEY and published by the Public School Publishing Company, Bloomington, Ill.

on which all pupils have written, be used for the analysis of handwriting. Every place in the paper where the handwriting is illegible should be clearly marked, after which all of the illegibilities should be marked on the chart. Actually, the chart is nothing but an error-analysis sheet, with all the common errors listed and with space for less common errors available. The chart is large enough for use with class groups of 40 pupils. Charting all of the errors in all papers gives data for the guidance of class as well as individual instruction.

The consistent use of this chart will increase the readability of handwriting in general, but it does not follow necessarily that improvement in handwriting will occur in specific details. The meaning of a word can be obtained from a few well-formed letters or from the general context. Close examination, such as is given by proofreaders to the printed page, must be given the samples of handwriting if improvement in specific details is to be obtained.

The supplementary data given on the chart will be of value to teachers interested in diagnosing the handwriting performance of their pupils.

*The West Chart for Diagnosing Elements of Handwriting.*¹—This chart provides for a complete objective diagnosis of certain elements found in handwriting. The author suggests that, although many teachers do now give attention to the handwriting of individual pupils, more would do so if they knew what to look for in the way of defects or how to provide the remedial treatment. The chart attempts to indicate symptoms of defective work, and the manual accompanying it gives helpful information, as to both diagnosis and remedial work.²

This chart may be used in diagnosing handwriting during the handwriting class or samples of handwriting obtained during other classwork. Its continuous use will lend expertness to diagnosis. The plan will fit into any program of instruction that is in use, adding to but in no way interfering with the values of that system.

¹ Devised by P. V. West and published by the Public School Publishing Company, Bloomington, Ill.

² P. V. WEST, Remedial and Follow-up Work, *Handwriting* (Elements of Diagnosis and Judgment of Handwriting), *Bulletin* 1, Bloomington, Ill.: Public School Publishing Company, 1926.

CHART FOR DIAGNOSING ELEMENTS OF HANDWRITING

By PAUL V. WEST

City.....Teacher.....

School.....Grade () Sec. ()

CHARACTERISTICS

SUGGESTED SYMBOLS

-
- 0 = No defects
 - 1 = Slight defect
 - 2 = Serious defect
 - (2) = Most serious defect

PUPIL		1 Rate (letters per min.)	2 Qual- ity		3 Letter Form		4 Slant			5 Coordi- nation		6 Motor Control			7 Unit Stroke						8 Spacing			9 Passage Between Words	Total
No.	NAME		Formal	Daily	No Habit	Wrong Habits	Extreme	Mixed	Changing	Wavering	Angular	Loose	Restrained	Mixed	Words			Letters							
															Wide	Narrow	Mixed	Wide	Narrow	Mixed					
			a.	b.	c.	a.	b.	c.																	
1																									
2																									
3																									
4																									
5																									
36																									
37																									
38																									
39																									
40																									

Just as the survey measures precede the proper use of diagnostic measurements of handwriting, so diagnosis should precede teaching to eradicate specific defects. Systematically, after periods of instruction, the teacher should apply survey measures to check improvement of the group and, possibly, of individuals. In handwriting, as in other school subjects, the test-teach-test technique will bring good results. It should be noted that both speed and quality of handwriting should improve with practice.

Readings for Further Study

1. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. XI. Boston: Houghton Mifflin Company, 1931.
2. W. S. MONROE, J. C. DEVOSS, and F. J. KELLY, *Educational Tests and Measurements* (rev. ed.). Chap. IV. Boston: Houghton Mifflin Company, 1924.
3. H. B. REED, *Psychology of Elementary School Subjects*. Chap. XIV.
4. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. XI.
5. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. III.

II. SPELLING

The ability to spell is important to all individuals but is concerned almost entirely with written work. Ordinarily, little use occurs for oral spelling in communication. The only reason that any individual must learn to spell words is to write them correctly in any record that he may wish to make. This is essential for the proper conveyance of ideas from one individual to another. The function to be developed in learning to spell is the ability to use the correct letters in their proper sequence in the given word.

Spelling is both highly objective and highly standardized. Classroom teachers in the past have generally been content to use the conventionalized column dictation test to measure the spelling achievement of their pupils. Because spelling words apart from meaningful context is an unnatural mental activity, and because it is monotonous to both pupils and teachers, efforts have been made to devise other types of instruments to measure spelling achievement. Regardless of what method is used, there must be valid and reliable tests to measure the effectiveness of that method.

Experiments with Spelling Tests.—Three investigators have used column (oral) dictation tests and sentence dictation spelling

tests.¹ The evidence indicates that these types of tests validate each other satisfactorily. There is some evidence that pupils spell more accurately on the oral dictation test than on the timed sentence test.

Guiler used three types of spelling tests in one investigation: a multiple-choice test with five responses, one spelled correctly and four misspelled, contained in sentences; a written recall test, in completion form, with the most common misspelling of the key word given; and an oral dictation test, in which the word was pronounced and then used in a sentence.² Phillips investigated the relative reliability of the oral dictation and two-response types of spelling tests at the college level.³ Again the oral dictation test ranked highest in one study, but in the second it was inferior to the two-response type of spelling test.

Epler⁴ and Lane⁵ devised 12 spelling tests, two each of six types of tests: oral dictation, sentence dictation, completion, two-answer and four-answer multiple-response, and true-false tests. Both investigators reported that the different types of tests were not equally reliable and valid, but their results do not agree as to excellence. Epler lists the types in order of excellence as (1) oral dictation test, (2) sentence dictation test, (3) four-answer multiple-response test, (4) completion test, (5) true-false test, and (6) two-answer multiple-response test. Lane reports the order of excellence of his tests as (1) oral dictation test, (2) four-answer multiple-response test, (3) sentence-dictation test, (4) two-answer multiple-response test, (5) completion test, and (6) true-false test.

¹ R. V. HUNKINS, *An Experiment in Column versus Dictation Spelling*, *Elementary School Journal*, 19: 689 (1919).

R. L. MORTON, *The Validity of Timed Sentence and Column Tests in Spelling*, *Journal of Educational Research*, 5: 444-447 (1922).

J. S. ORLEANS, *The Ability to Spell*, *School and Society*, 23: 407-408 (1926).

² W. S. GUILER, *Validation of Methods of Testing Spelling*, *Journal of Educational Research*, 20: 181-189 (1929).

³ D. P. PHILLIPS, *Comparison of the Two-response and Dictated Recall Types of Spelling Tests*, *Journal of Educational Research*, 23: 17-24 (1931).

⁴ NORA EPLER, *A Study of Objective, Teacher-made Spelling Tests*, unpublished Master's thesis, University of Southern California, 1933.

⁵ J. B. LANE, *A Study of Teacher-made, Objective Spelling Tests*, unpublished Master's thesis, University of Southern California, 1934.

Two studies have been made in which both teacher-made objective tests and standardized spelling tests have been used.¹ Hight's results indicate that teachers can devise objective spelling tests that are possessed of characteristics similar to those of standardized tests. The writer, with Louise E. Weller, conducted an investigation involving six types of spelling tests: (1) a two-response, written recall spelling test, (2) a two-answer multiple-response written recall spelling test,² (3) a four-answer multiple-response written recall spelling test,³ (4) a completion spelling test, (5) a column dictation spelling test, and (6) the New Stanford Achievement Test, Dictation Exercise.⁴

The two-response written recall test was presented in mimeographed form to the pupils, as follows:

Directions: In this test, some of the words are spelled correctly, and some are spelled wrongly. Look at each word carefully; if it is right, place a check in the space before the word and under R; if wrong, a check in the space after the word and under W.

Example:

R		W
_____	verry	x
x	brave	_____

The completion spelling test was presented in mimeographed form to the pupils, as follows:

Directions: In this test, the first letter or the first two letters of a word are given. Fill in the blanks with the letters that are not given, in order to finish spelling the word. After each blank is a definition or another word that means the same thing as the word you are to spell.

Example:

b_____ (a place to sleep on, a couch)
 h_____ (a place in which to live, an abode)

¹ See data from Millicent O. Hight, Chap. V (pp. 95-98).

² This test was similar in form to the spelling test included as Part I of the Cross English Test, devised by E. A. Cross and published by the World Book Company, Yonkers-on-Hudson, N. Y.

³ This test was similar to the spelling test included as Part I of the Columbia Research Bureau English Test, devised by H. R. Steeves, A. Abbott, and B. D. Wood and published by the World Book Company, Yonkers-on-Hudson, New York.

⁴ Devised by T. L. Kelley, G. M. Ruch, and L. M. Terman and published by the World Book Company, Yonkers-on-Hudson, N. Y.

The two-answer and four-answer multiple-response spelling tests also were presented to the pupils in mimeographed form, but the column dictation and New Stanford sentence dictation spelling tests were presented orally to the pupils.

The validity coefficients obtained in this investigation are presented in Table XVII.

TABLE XVII.—MEASURES OF RELATIONSHIP FOR SIX SERIES OF SPELLING SCORES¹
N = 174

Tests	Two-answer multiple-response	Four-answer multiple-response	Completion	Column dictation	Sentence dictation
Two-response.....	0.927	0.778	0.759	0.351	0.262
Two-answer multiple-response.....	0.832	0.797	0.496	0.308
Four-answer multiple-response.....	0.784	0.425	0.311
Completion.....	0.193	0.497
Column dictation.....	0.823

¹ LOUISE E. WELLER and M. E. BROOM, A Study of the Validity of Six Types of Spelling Tests, *School and Society*, 40: 103-104 (1934).

Since all of the coefficients in Table XVII are positive, all of the tests measure the same thing or things to some extent. The first four tests, the two-response test, the two-answer and four-answer multiple-response tests, and the completion test, validate each other very well, and yet these four tests show low validity with the oral column dictation test and the sentence dictation test, which were found to be the best measures of spelling achievement in earlier studies. The last two tests validate each other very well also. Apparently, the last two tests measure something other than is measured by the first four tests, perhaps vocabulary skills and proofreading abilities. Previous studies were conducted in upper elementary-school and high-school grades (except one study at the college level). The data in this investigation were obtained in grades 2 to 5 inclusive, and this also may have had some effect on the outcome.

In general, the oral column dictation test and the sentence dictation test seem to be the best measures of spelling achieve-

ment so far devised. The evidence from numerous studies indicates that other types of spelling tests show varying degrees of reliability and validity, without positive evidence that they measure spelling ability only. There is some evidence that proofreading and vocabulary skills may interfere with the measurement of spelling ability in certain types of objective tests other than the oral column and sentence dictation tests.

Standardized Materials for the Measurement of Spelling Achievement.—Many studies have been made dealing with the proper spelling vocabulary, and the findings clearly show that the average writing vocabulary is of small compass, including relatively simple words. The typical spelling list for use in progressive elementary schools includes now from 3,000 to 4,000 words, with the meaning of which the pupils are familiar.

Such spelling vocabularies have been scaled as to their difficulty to pupils in different elementary grades in so far as correct spelling is concerned. The words, when grouped as to difficulty in spelling, constitute the spelling scale. Spelling scales actually are only groups of words arranged and grouped according to difficulty. Their major purpose is to provide many words from among which certain ones can be chosen for incorporation in spelling tests.

The Ayres Spelling Scale.—The original Ayres scale included 1,000 words, including those which occurred at least forty-four times in 368,000 words written by 2,500 different persons. The words in this group may be thought of as including the fundamental English words in common use in written communication of ideas. One advantage is that they were taken from materials written by adults, so that ability to spell them insures the proper training of school pupils for active social participation in adult life.

After selecting the 1,000 words, Ayres proceeded to determine the difficulty of each one. They were divided into 50 lists of 20 words each. They then were spelled by 70,000 elementary pupils in 84 school systems scattered throughout the United States. The total number of spellings of all words was 1,400,000, an average of 1,400 spellings per word. When grouped according to their difficulty in 26 lists, the words formed the Ayres Spelling Scale. Later, Buckingham added an additional 505 words, chosen according to agreements among spelling books and dis-

tributed according to their spelling difficulty, forming the Buckingham Extension of the Ayres Spelling Scale.¹

Buckingham writes concerning the extension of the scale:²

This scale is not a test but a list of words from which the teacher may make a test. . . . All the words in each column are of approximately equal spelling difficulty. The steps in spelling difficulty from each column to the next are approximately equal steps. The numbers at the top indicate about what percent of correct spellings may be expected among the children of the different grades, at the end of the school year. For example, if 20 words from column H are given as a spelling test it may be expected that the average score for an entire second grade spelling them will be about 79 percent. For a third grade it should be about 92 percent, for a fourth grade about 98 per cent, and for a fifth grade about 100 percent. . . . Since the per cents given in each column are for the end of the school year, the percents are the upper limits for the given grades.

The words in the Buckingham Extension of the Ayres Spelling Scale are arranged in 32 lists. An excerpt from the scale is shown on page 164. It should be noted that the words in italics are those added by Buckingham.

Twenty words, the number ordinarily used in tests chosen from the scale, are enough to secure a reasonably reliable measurement of the spelling achievement of a group, but this number is inadequate for a valid and reliable measurement of an individual's spelling achievement. From 50 to 100 words should be used in a test of an individual's spelling ability.

*The Iowa Spelling Scales.*³—These scales include 3,000 of the words most frequently used in adult correspondence, as determined by W. N. Anderson. Ashbaugh used the frequency arrangement established by Anderson in selecting the 3,000 words. The words were divided into three lists of 1,000 each; and these, in turn, were subdivided into 10 lists of 100 words each. These lists were sent to a large number of schools in Iowa, and all of the words were given to all of the pupils in all of the grades from grade 2 through grade 8 inclusive. In grades 2

¹ Devised by B. R. Buckingham and published by the Public School Publishing Company, Bloomington, Ill.

² B. R. BUCKINGHAM, *Extension of the Ayres Spelling Scale* (new ed.) p. 2.

³ Devised by E. J. Ashbaugh and published by the Public School Publishing Company, Bloomington, Ill.

164 MEASUREMENTS IN THE ELEMENTARY SCHOOL

	G	H	I	J	K		
Second Grade→	84	79	73	66	58		
Third Grade→	94	92	88	84	79		
Fourth Grade→	99	98	96	94	92		
Fifth Grade→	100		99	98	96		
			Sixth Grade→	100	99		
by have are had over must make school street say come hand ring live kill late let big mother three land cold hot hat child ice play sea <i>bread</i> <i>come</i> <i>eats</i> <i>food</i>	day eat sit lot box belong door yes low soft stand yard bring tell five ball law ask just way get home much call long love then house year to I as send one has some if how her them other baby well about men for ran was that his led lay	nine face miss ride tree sick got north white spent foot blow block spring river plant cut song winter stone free lake page nice end fall feet went back away paper put each soon came Sunday show Monday yet find give new letter take Mr. after thing what its very or	thank dear west sold told best form far gave alike add <i>brave</i> <i>corn</i> <i>dance</i> <i>dinner</i> <i>egg</i> <i>looks</i> <i>rich</i>	seven forget happy noon think sister cast card south deep inside blue post town stay grand outside dark band game boat rest east son help hard race cover fire age gold read fine cannot May line left ship train saw pay large near down why bill want girl part still place report	never found side kind life here car word every under most made said work our more when from wind print air fill along lost name room hope same glad with mine <i>chair</i> <i>forgot</i> <i>hang</i> <i>meal</i> <i>mouse</i> <i>sits</i> <i>store</i> <i>supper</i>	became brother rain keep start mail eye glass party upon two they would any could should city only where week first sent mile seem even without afternoon Friday hour wife state July head story open short lady reach better water round cost price become class horse care try move delay pound behind	around burn camp bear clear clean spell poor finish hurt maybe across tonight tenth sir these club seen felt full fail set stamp light coming cent night pass shut easy <i>birds</i> <i>bone</i> <i>cloud</i> <i>garden</i> <i>goose</i> <i>knife</i> <i>mouth</i> <i>oak</i> <i>peach</i> <i>pole</i> <i>queen</i> <i>rope</i> <i>season</i> <i>space</i> <i>stands</i> <i>wagon</i> <i>wheat</i> <i>window</i>

through 4, the pupils spelled 50 words per day; and in grades 5 through 8, 100 per day. All the spellings were made without preliminary study or help from the teacher.

The total number of spellings in the seven grades was 4,662,200; and at least 200 spellings were used for every word. The difficulty of every word for pupils in each of the seven grades was definitely determined statistically.

The words finally used in the scale, 2,977, were arranged in seven groups for use in grades 2 through 8, one scale for each grade level. Each list is divided into steps, according to difficulty. There are several levels of difficulty for every grade. A number of words are displaced one step, and a few two or more steps, owing to consolidation. Since the median increase in difficulty from grade to grade is about two and one-half steps, some words may be displaced by as much as an entire grade.

There is provision for repetition of words from grade to grade. A separate scale difficulty is given in such instances. Thus the word "about," which occurs in Step 9 of the Grade 3 Scale in the group of words that should be spelled correctly by 81 per cent of the pupils in the grade, also is found in the Grade 4 Scale in Step 5 in the group of words which should be spelled correctly by 95 per cent of the pupils in grade 4. The percent-of-correctness norms given in the scales are based upon testing done at the middle of the school year.

A part of the Grade 4 Scale is shown on page 166.

This illustration serves to make clear the arrangement of the scale according to the difficulty of the words, both by steps and within the steps of the scale. The number of words is adequate both for repeated testing and for testing the spelling ability of individual pupils. The extensiveness of the spelling material in the scale, plus the careful standardization of these materials, makes the scale probably the most valuable measuring device now available to classroom teachers.

Standardized Tests in Spelling.—A classroom teacher may select words from a standardized spelling scale and so arrange many tests for which norms are available. Numerous standardized tests are available in this subject. These are usually oral column dictation tests or sentence dictation tests.

		GRADE IV	
STEP 1—100%	in	your	hat
all	into		head
at	last	96%	Jan.
did	let		longer
dog	letter	an	pay
me	may	cake	place
thank	mother	cent	poor
thing	must	city	ring
to-day	not	corn	river
	old	down	same
	on	five	send
STEP 2—99%	one	food	some
and	over	free	summer
are	say	going	sweet
big	she	gold	table
book	that	hard	then
but	think	her	time
day	this	how	to-night
dear	way	live	us
good	when	looking	wish
have	will	man	work
he	with	my	working
his		read	
is	STEP 4—96%	reader	94%
it	97%	red	
land		sand	am
like	apple	south	bad
little	bell	tell	bed
look	butter	ten	bee
May	by	we	cannot
out	cook	STEP 5—94%	coat
run	door	95%	end
see	eat		fast
ship	green	about	faster
tree	ice	after	Feb.
up	if	age	for
was	left	ago	has
what	love	baby	him
year	made	bank	just
you	make	bill	kind
	milk	box	king
STEP 3—98%	night	dark	Monday
as	said	date	Nov.
away	school	dinner	of
be	six	do	papa
boy	snow	doing	pig
can	state	east	play
come	Sunday	ever	real
each	them	face	ride
get	they	foot	road
give	three	from	Sat.
go	top	frost	seeing
got	under	gave	silk
hand	why	had	stove
	wind		sun

*The Iowa Spelling Test.*¹—This test includes two groups of spelling words, the first group of 25 words being given in sentence dictation form and the second of 25 in oral or column dictation form. The words are taken from the Iowa Spelling Scales.

TEST 2

FOR GRADES FIVE AND SIX

Dictation Exercise

Start with

watch at

- 60 The *captain* was a *capable* man.
 26 We are to *entertain* a *guest* to-night.
 58 A strange *animal* *appeared* before us.
 30 The *stories* were *written* by the class.
 4 We are *enjoying* the *climate* in California.
 42 The *soldier* was *crippled* in the *worst* battle.
 23 He was *manager* and *treasurer* of the *local* club.
 5 A *correct* *account* of the *recent* *lecture* was given.
 49 An *improvement* was made in the *machine* last year.
 33 The *author* was *satisfied* with the *publication* of his book.

Stop—33

Now have the children write the following words on the blank side of the paper. Pronounce these as you would an ordinary spelling lesson.

Spelling List

barrel	parties	suppose	doubt	maintain
bigger	permit	uniform	easily	receive
explained	quiet	voting	federal	struggle
grip	ribbon	accept	increased	signed
include	silence	courage	laid	republican

In the sentence dictation test, the sentences are read once only, and pupils are instructed to listen to the entire sentence before writing. This is a timed test. The column dictation test is not timed. Tentative norms for the two parts of the test are given separately, in terms of number of words correctly spelled and also in terms of average percentage of spelling accuracy. These

¹ Devised by H. A. Greene and published by the Extension Division, University of Iowa, Iowa City.

norms were established on midyear performance, and they are based on an average of 170 pupils per school grade. The test author indicates that the test is to be given without previous teaching and that the directions accompanying the test are to be followed exactly. Separate tests are available for grades 3 and 4 (Test I), 5 and 6 (Test II), and 7 and 8 (Test III). The brevity of the sentence and column dictation tests makes both poor measures of the spelling ability of individual pupils. The scores should be combined, or two forms of the test should be used when teachers are attempting to secure valid measures of the spelling achievement of individual pupils.

Column Dictation Tests.—There are numerous tests of this type available to classroom teachers. Certain of them are individual spelling tests, and others are parts of batteries of school achievement tests. Typical instruments will be treated briefly.

The Spelling Test of the Unit Scales of Attainment¹ is a list of spelling words to be pronounced individually to pupils. The words may be pronounced twice, if necessary. In some instances, where two words have the same pronunciation, a short sentence or phrase may be used to convey the correct meaning. In general, the procedure is similar to that used in the usual classroom tests in spelling.

The Spelling Test of the Progressive Achievement Tests² requires that words be pronounced, then used in a sentence, and then pronounced a second time, after which the pupils write the word only in the test booklet. One form of the three available in the Elementary Battery of this test is given on page 169.

It should perhaps be noted that two forms of this test should be dictated when a classroom teacher is attempting to measure the spelling achievement of individuals. The single form, however, is probably sufficiently long to permit the adequate measurement of spelling achievement of groups of pupil.

The Guy Spelling Scales³ include three tests of the type discussed above, each 50 words in length. The tests are intended

¹ Devised by L. J. Brueckner, W. A. Anderson, and M. J. Van Wagenen and published by the Educational Test Bureau, Inc., Minneapolis, Minn.

² Devised by E. W. Tiegs and W. W. Clark and published by the California Test Bureau, Los Angeles.

³ Devised by J. F. Guy and published by the Public School Publishing Company, Bloomington, Ill.

for use in grades H 2 through 9. The first 30 words are intended for use in grade 2; the first 35, in grade 3; the first 40, in grade 4, the first 45, in grade 5; and the entire 50, in grades 6 through 9. Norms based on 48,400 spellings are available. The wide grade spread and the relative shortness of each test indicate that the measures of spelling achievement of individuals will be inaccurate, at least in the lower grades.

FORM A

1. first	She will be FIRST .	first
2. any	Have you ANY pencils?	any
3. glass	We see thru GLASS .	glass
4. above	The stars are ABOVE us.	above
5. turn	It is your TURN .	turn
6. lesson	I know my LESSON .	lesson
7. change	We shall CHANGE the tire.	change
8. account	We must go home on ACCOUNT of the rain.	account
9. beside	We sat BESIDE the fire.	beside
10. need	I NEED your help.	need
11. young	First grade children are YOUNG .	young
12. expect	I EXPECT to see her soon.	expect
13. none	You have a pencil, I have NONE .	none
14. second	You will be SECOND at bat.	second
15. duty	I know my DUTY .	duty
16. since	He has not been here SINCE last week.	since
17. though	We lost the game, even THOUGH we played hard.	though
18. does	He DOES well in spelling.	does
19. field	The farmers are in the FIELD	field
20. allow	Please ALLOW a wide margin.	allow
21. purpose	His PURPOSE is to read a book.	purpose
22. together	We shall sing TOGETHER .	together
23. pleasant	This is a PLEASANT day.	pleasant
24. certain	I am CERTAIN I can go.	certain
25. really	Did you REALLY hear that?	really
26. decide	I cannot DECIDE what to do.	decide
27. receive	You will RECEIVE a letter.	receive
28. mere	It is a MERE trifle.	mere
29. reference	This is a good REFERENCE book.	reference
30. proceed	We shall PROCEED with our lessons.	proceed

The Spelling Test of the Metropolitan Achievement Tests¹ is similar to those previously mentioned, except that it is staggered in arrangement. For example, the Intermediate Battery, intended for use in grades 4, 5, and 6, includes a spelling test for

¹ Devised by R. D. Allen, H. H. Bixler, W. L. Connor, and F. B. Graham and published by the World Book Company, Yonkers-on-Hudson, N. Y.

pupils in grade 4 which includes only words 1 to 45. Pupils in grade 5 spell words 6 to 55 inclusive; in grades 6, words 16 to 65 inclusive.

The Morrison-McCall Spelling Scale¹ includes eight different and duplicate tests. The lists are equally difficult. Extensive norms are available for these tests. The scale is useful in grades 2 through 9 for group measurement purposes, but, like the Guy Spelling Scales, its findings should be questioned for individuals unless two or more forms have been used. It is difficult to secure accurate measurements of achievement of individuals with a 50-word spelling test in eight grades.

Sentence Dictation Tests.—This form has been illustrated on page 167 in the presentation of the Iowa Spelling Test. In general, the form of tests of this type is similar.

The Spelling Test of the Public School Achievement Tests² includes 114 words arranged in 27 sentences, in Forms 3 and 4; and 122 words arranged in 31 sentences, in Forms 1 and 2. Not all of the words are dictated to pupils in all grades; the arrangement of the sentences is staggered in a fashion similar to that of the Spelling Test of the Metropolitan Achievement Tests.

The form and content of this test are shown by the partial examination given below.

SENTENCES TO BE DICTATED, FORM 4

	(4)	1. <i>Play the game for me.</i>	} Dictate sentences 1 to 18 in grades 3 and 4
	(7)	2. <i>See the fat pig.</i>	
	(11)	3. <i>The little cat ran away.</i>	
	(14)	4. <i>He is looking over the hill.</i>	
	(18)	5. <i>Please take your coat.</i>	
Dictate sentences 8 to 27 in grades 5, 6, 7, and 8	(103)	24. <i>The planning of the exhibit was original and thorough.</i>	
	(107)	25. <i>The grammar assignment was based on analysis.</i>	
	(111)	26. <i>Unnecessary communication frequently brings permanent injury.</i>	
	(114)	27. <i>The politician changed his plans definitely and materially.</i>	

This test is masked as a dictation exercise; pupils are not told that it is actually a spelling test. The words to be checked for

¹ Devised by J. C. Morrison and W. A. McCall and published by the World Book Company, Yonkers-on-Hudson, N. Y.

² Devised by J. S. Orleans and T. L. Torgerson and published by the Public School Publishing Company, Bloomington, Ill.

correctness of spelling are those in italics. Only the italicized words are scored, and misspellings of other words are disregarded. This is one fault of this type of test, since disregarding the incorrect spelling of any word may tend to fix the inaccuracy in the pupil's mind.

The test is not timed, but pupils who delay the class may be urged to write more rapidly. Penalties are not placed against pupils because of poor handwriting.

The vertical lines in each sentence are for the purpose of dividing it into short groups of words to aid the examiner in uniformity of pronunciation. The numbers in the parentheses at the left of the sentences indicate the cumulative number of words in the test. For example, the number 18 appears before sentence 5 in the test used as an illustration. This means that in sentences 1 through 5 there are 18 scorable words.

This is a superior test. The words are taken from the Iowa Spelling Scales, insuring the validity of the test materials; the length of each form of the test insures accuracy of measurement. Norms are available.

The Dictation Test of the New Stanford Achievement Test¹ is similar in construction to the Spelling Test of the Public School Achievement Tests. It does not have the number of scorable words in the different sentences indicated. It is staggered similarly to the form of the Spelling Test of the Metropolitan Achievement Tests.

The 108 words included in the test are arranged with several words in each sentence. These words, excepting some few of the more difficult ones in the portion of the test for use in the upper elementary grades, are taken from standardized spelling scales.

This test is a revision of the Dictation Test of the Stanford Achievement Test by the same authors, a pioneer instrument of its type. The present test is well constructed, and it has been carefully standardized. The measures that it yields are reliable and valid.

In concluding the discussion of instruments for measuring achievement in spelling, it should be pointed out that the standardized tests are restricted almost entirely to the two types of tests shown by research findings to be the most reliable and valid

¹ Devised by T. L. Kelley, G. M. Ruch, and L. M. Terman and published by the World Book Company, Yonkers-on-Hudson, N. Y.

measures of ability to spell: column and sentence dictation tests. A considerable amount of work has been done, both in developing spelling vocabularies and in devising spelling scales and tests, and classroom teachers should familiarize themselves with the products of this research.

Obviously, it is not enough to know merely the score representing a pupil's achievement on a standardized spelling test. Such measures indicate only a general level of spelling achievement, either for an individual or for the class group. Classroom teachers should seek to ascertain the causes of difficulties in spelling, whether poor study habits, a lack of knowledge of word meanings, carelessness in omission of letters, phonetic spelling, or the like are causing errors. After the diagnosis has been made, the teacher should utilize remedial practices that have been found to be valuable in improving spelling achievement. Many suggestions of merit may be found in the references listed below.

Readings for Further Study

1. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. X.
2. W. S. MONROE, J. C. DEVOSS, and F. J. KELLY, *Educational Tests and Measurements* (rev. ed.). Chap. V.
3. E. M. PAULU, *Diagnostic Testing and Remedial Teaching*. Chap. VI. Boston: D. C. Heath and Company, 1924.
4. H. B. REED, *Psychology of Elementary School Subjects* (rev. ed.). Chaps. XII, XIII.
5. M. R. TRABUE, *Measuring Results in Education*. Chaps. VI, VII.
6. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. X.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chaps. II, IV.

III. COMPOSITION, MECHANICS, GRAMMAR

The major objective in oral and written expression is the development of the ability to communicate ideas effectively. Since it is difficult, if not impossible, to control speech sufficiently to apply standardized techniques in its evaluation, the discussion following will be limited to written language.

It has been emphasized in the preceding discussion that handwriting and spelling are essential in written communication. These, however, are not more important than is the ability to use

mechanics such as capitalization and punctuation accurately or the ability to express ideas clearly and correctly in sentences and paragraphs. Unless an individual uses these abilities properly, the thought that he is attempting to express may be obscured. For example, a sentence that is made up of words that do not say definitely and precisely what is meant is wasted, and such a sentence may even be a contributing factor in causing a failure of comprehension of the thought of the communication. Similarly, an improper use of punctuation or the entire lack of punctuation in a sentence or a paragraph may reduce the words to a conglomerate mass from which it is difficult to sort and group the words properly to convey meaning to a reader.

Two types of instruments are available for the measurement of achievement in written expression: composition scales and tests of mechanics and grammar.

Composition Scales.—These instruments are scoring instruments, not tests. The lower limits of most of these scales show weak thought content and numerous errors in grammar and mechanics, whereas the upper limits are almost entirely free from such errors.

Composition scales ordinarily are derived on the basis of an average or combined judgment of experts, so that they are not entirely objective in derivation, as are standardized tests. Since, in use, they are compared by classroom teachers to pupil-written themes, until the pupil-written theme in each instance is located on the scale in terms of similarity to a sample theme in the scale, composition scales are somewhat subjective in use also. It is obvious that it is impossible to measure English composition, which is influenced by many subtle factors, in an objective way when comparison only is utilized.

The most troublesome differences in judging pupil-written themes in terms of a composition scale come in themes that are excellent in mechanical and grammatical matters but thin or stupid in thought content and those which are excellent in thought content but bad in matters of form. Many of the scales provide little help in this situation. They have for their field the identification and evaluation of general merit in written expression, and they provide no means of analyzing specifically such merit.

The teacher who uses composition scales must exercise very careful matching of her pupil-written themes against the examples in the scale. The scale will have fulfilled its function when the classroom teacher finds that her own judgments have become reasonably well standardized in terms of scale values. When a classroom teacher is able to use a composition scale accurately and readily, she is reasonably certain of achieving results in marking composition materials when not using a scale that are equivalent to the markings that she would obtain when using the scale. It is obvious that composition scales have their greatest usefulness in aiding inexperienced teachers to establish standards of judgment as to the merit and deficiencies of pupil-written compositions.

*The Hudelson English Composition Scale.*¹—This is a scale widely used for general purposes in evaluating composition quality. It includes 16 samples of written composition, scaled at uniform distances. This tends to avoid confusion in the use of the instrument. Only two specimens, the upper two, were not written by school children.

The scale provides instructions for obtaining pupil-written themes and a time limit, 30 minutes, during which the compositions must be written. When scoring, the teacher reads a theme and then compares it with the scale samples. "This is better than 2.0," he says to himself, "better than 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0. It is not so good as 5.5, but it is nearer in merit to 5.5 than it is to 5.0; so I shall give it the scale value of 5.5."²

The ratings represent steps above zero merit or no composition merit at all. The specimen rated 5.0 has 25 per cent more merit than the one rated 4.0, and sample 6.0 represents exactly twice as much composition merit as does sample 3.0. Composite median norms as of midyear are available for all elementary school grades above grade 4.

Two samples from the scale are given below. The second has just twice the composition merit of the first. According to the norms, the first represents midyear median composition merit of below grade 4, whereas the second represents midyear median composition merit of grade 6.

¹ Devised by Earl Hudelson and published by the World Book Company, Yonkers-on-Hudson, N. Y.

² EARL HUDELSON, *Hudelson English Composition Scale* (1st rev.), p. 1.

*Sample 1**Value 2.0*

The Most Exciting ride I ever had.

The Most exciting ride I ever had was a Hay ride, it was early in the morning when we went out on the hay ride it was quite a injoyable trip every one seemed to be so cheerfully the rode that we were traveling on it was very hilly on of the parties took sick and far a little while no one did not think that the Girl were as sick as she was all at once she come mence comeplaining so she arroused ones curosimy we found out that the girl were verry ell thought she was going to die.

*Sample 5**Value 4.0*

The Most exciting Ride I ever had.

It was just after a very hard Rain. and the Roads were slippery and muddy. My old friend and I honomed a car and started out for a spin. We had'nt gone far before we saw that the Road swere in no condition for motoring. We kept going thought and never stopped for nothing that norning. Along about the middle of the evening wa saw just ahead of us a lantty slaping hill. We though we could make it all right so we started though when we reached the middle of the clay hill we could'nt go any further we tried and tried to get out of the place but all in vain. So we had to get down and go all the way home on foot. We sent for car the next day and some one had took all the tires we had on the car. We soon had her right again and ham ben ridding on her sence then.

The advantage of this scale lies not only in its arrangement but also in the training materials that it provides to teachers. Thirty sample compositions, arranged in three series of 10 each, are available in the scale for training purposes. They have merit values also included in the scale, so that inexperienced teachers may rate them and then compare their own judgments with that of experts. Furthermore, they may be studied to determine why the experts ranked them as they did. If teachers use the training materials accurately and extensively before attempting to judge pupil-written themes, the judgments will be more reliable and valid than those obtained without previous training.

*The Willing Scale for Measuring Written Composition.*¹—This scale includes eight sample compositions, arranged arbitrarily on a scale at equal intervals. The values of the different samples occur at intervals of 10 points, with the poorest sample rated 20 and the best rated 90. The scale value 20 means from 15.0 to

¹ Devised by M. H. Willing and published by the Public School Publishing Company, Bloomington, Ill.

24.9; 30 means from 25.0 to 34.9; etc. A sample of the scale appears below.

50

One time mother and father were going to take sister and I for a long ride thanksgiving. We had to go 60 miles to get there, When sister and I herd about it we were very glad. It was a very cold trip. We four all went in a one seated automobile. Dady drove and mother held me and sister sat on the top the top was down. Mother could not hold sister for she was two heavy. When we got there they had a hot fire ready for us and a goose dinner. We were there over night. In the morning it was hot out. This was on a farm. Sister and I got to go horseback riding. It was lots of funs. They had children. The children were very nice. Our trip home was very cold. When we got home it had snod.

Number of mistakes in spelling, punctuation, and syntax per hundred words, 14.

This scale attempts to provide measures both of story value and of form value. Compositions are to be written in 20 minutes, with five minutes thereafter to finish the writing, make corrections, and count the number of words written. The completed pupil-written themes are rated first for story value, by the usual process of comparing them to the samples in the scale. After all specimens have been judged, the judgments are rechecked, after which each composition is marked with the value of the scale composition that it most nearly resembles. When these judgments have been made, the scorer marks carefully all errors in grammar, punctuation, capitalization, and spelling. The total number of errors is then multiplied by 100, and the result is divided by the number of words in the composition. The quotient is the index of the form value. Thus, each composition is given two separate scores. Norms for both score series are available on the scale for grades 4 through 8.

Uses of Composition Scales.—Composition scales have their greatest value in establishing the theme writing of groups. In attempting to evaluate the achievement of individuals in written composition, expert rating skill is necessary as well as a series of written compositions. It is obvious that a classroom teacher can learn much more about the achievement of a class group than she can about that of any individual when only one judgment is given.

By measuring compositions written by the same group of pupils at different times, a teacher can evaluate the amount of improvement that results from teaching during intervals between the testings. If different methods of teaching have been used during these intervals, the teacher is able to evaluate the comparative efficiency of these methods and so guide her teaching procedures to obtain maximum results. Finally, by comparing the achievements of her pupils against the norms, the teacher is able to determine whether or not her pupils are writing as well as they should write.

She should not forget the subjective character of the measures yielded by the scales, and she should avoid thinking of these measures as being as precise and accurate as scores yielded by standardized tests. Avoiding this source of error will tend to make the scales of greater value for classroom teaching purposes.

Standardized Tests.—The types of content commonly found in standardized tests measuring mechanics and grammar in written expression are well illustrated in the batteries of achievement tests.¹ In the Metropolitan Achievement Tests for grades 2 and 3, there is a language usage test. The battery for grades 4, 5, and 6 includes tests of language usage and of capitalization and punctuation; and for grades 7 and 8, language usage, capitalization and punctuation, and grammar. The language tests of the Unit Scales of Attainment for the intermediate and upper grades measure punctuation, capitalization, and language usage. The difficulty of the content differs, increasing for the upper grades. The language tests of the Public School Achievement Tests measure language usage in grades 3 through 8 and grammar in grades 6 through 8. A test of language usage is included in the New Stanford Achievement Test, for use in grades 2 through 8. The Progressive Achievement Tests include a test of capitalization and punctuation in the primary grades 1, 2, and 3; tests of capitalization and punctuation and language usage in the intermediate grades 4, 5, and 6; and tests of capitalization and punctuation, language usage, and grammar in the upper grades 7 and 8. The discussion of typical standardized tests follows.

Standardized Tests. Grammar.—During recent decades, there has been a change in the teaching of grammar. Grammar is

¹ These have been cited previously in the section dealing with spelling, pp. 168-171.

seldom taught in modern elementary schools as an end in itself but rather as a means to an end. Formal grammar has been largely replaced by functional grammar, which attempts to make clear to pupils the nature of errors, thus providing a basis for self-checking and self-correction. The standardized tests reflect this trend in teaching.

*The Franseen Diagnostic Tests in Language.*¹—This is a group of three tests measuring (1) pronouns, (2) verbs, and (3) varied forms. The test on pronouns includes 120 items, grouped in 14 divisions, each of which deals with a single type of construction. In turn, each item within any group represents a specific instance of that construction. All of the exercises are similar in form, simplifying the mechanical difficulties of the test for pupils responding to it. The form and content of the test may be illustrated by an excerpt from the test on pronouns.

GROUP F

- | | | |
|------------------------|------------|--------------|
| 1. The guard commanded | I | } to stop. |
| | me | |
| 2. They were expecting | him | } to arrive. |
| | he | |
| 3. What are you asking | her | } to do? |
| | she | |
| 4. Father won't permit | we | } to stay. |
| | us | |
| 5. Hurry fast and tell | they | } to wait. |
| | them | |
| 6. Your sister told me | whom | } to see. |
| | who | |

SCORE.....

This test does not provide norms data, but it has directions for both pupil and class diagnosis, both gross and specific, and suggestions for remedial teaching. The score sheets for each type of test include statements of the meaning of each group of items in the test. For example, the section illustrated, Group F of the pronouns test, tests the use of the nominative instead of the objective form as the subject of an infinitive.

The Gregory Diagnostic Tests in Language² is similar in content to the Franseen tests but shorter. The pronouns section

¹ Devised by C. E. Franseen and published by the C. A. Gregory Company, Cincinnati, Ohio.

² Devised by Helen Gregory, M. A. and C. A. Gregory and published by the C. A. Gregory Company, Cincinnati, Ohio.

includes 55 items; the verbs section, 60 items; and the miscellaneous constructions section, 41 items dealing with adverbs, nouns, prepositions, conjunctions, and adjectives. Norms are available for this test. The Franseen test was devised for use in grades 3 through 8; and the Gregory, for use in grades 4 through 9.

*The Grammar Test of the Metropolitan Achievement Tests.*¹—This is typical of the short survey tests of grammar found in the batteries of school achievement tests and in certain series of language tests. It is not intended to give a comprehensive picture of the grammatical achievement of pupils; nor does it attempt to yield a diagnosis of pupil weaknesses in grammar. It is useful primarily in determining roughly the status of class groups, and it precedes the use of diagnostic instruments. The form and content of the test is indicated in the excerpt below.

Directions. Read this paragraph. Then answer the questions below it.

The goal, the aim, of our schools will be reached when every boy and girl who is of school age is able to obtain an education. We have not yet achieved all that is desired, but we have made marked progress. In 1870 only 57 per cent of the children of school age were in school. The increase from 1870 to 1900 was greater than for any previous similar period.

74. How many sentences are there in the above paragraph? () 74
 75. Write the first and last words of a simple sentence. . . . () 75
 76. Write the first and last words of a compound sentence. () 76
 77. Write the first and last words of a complex sentence. . . () 77

In the parentheses after each of the following write the number of one word or expression from the paragraph above that it describes.

78. a conjunction. . . . () 81. a verb in the future tense. ()
 79. a relative pronoun. () 82. the comparative degree of an adjective ()
 80. a noun in apposition () 83. the antecedent of a relative pronoun. ()

Capitalization and Punctuation Tests.—The Seaton-Pressey Diagnostic Tests in English Composition² provide separate survey tests of capitalization and of punctuation. The first excerpt on page 180 is from the test of minimal essentials of capitalization; the second, from the test of minimal essentials of punctuation.

¹ Devised by Helen Gregory, M. A., and C. A. Gregory, and published by the C. A. Gregory Company, Cincinnati, Ohio.

² Devised by J. T. Seaton and S. L. Pressey and published by the Public School Publishing Company, Bloomington, Ill.

1. We are looking for mary evans and helen davis.
2. I am reading about president lincoln and general grant.
3. Mr. and mrs. james t. brown are visiting dr. s. w. jones.
4. He wants to be a doctor or a judge.

1. Who has seen my hat It is an old felt hat.
2. Mr and Mrs Jones are not at home.
3. Fred was absent Monday Tuesday Wednesday and Friday.
4. They live at 134 Wood Street Hamilton Ohio.
5. Helens hair is brown.

These tests yield satisfactory gross class measures of achievement in capitalization and punctuation. If two forms are used, the results will give a satisfactory individual measurement. Similar but more difficult and more extensive tests are included in the Pressey Diagnostic Tests in English Composition¹ intended for use in grades 6 through 8 in the elementary school.

The Pressey-Seaton tests include also tests of language usage and of sentence structure. The Pressey tests include a grammar test, which is essentially a language usage test, and a sentence structure test. These tests are similar in form to those illustrated above. It should be noted that the mental operations required in responding to these tests are of the nature of proof-reading. There is no question that ability to write adequately and accurately and proofreading abilities are entirely distinct. These tests provide a basis for pupil self-correction, but high achievement on these tests does not guarantee good written expression.

Other Tests.—One interesting test that combines the measurement of capitalization and punctuation is the Leonard Diagnostic Test in Punctuation and Capitalization.² This should be useful in grades 5 through 8. The Los Angeles Diagnostic Test in Language³ provides measures of capitalization, punctuation, and language usage. This is a survey instrument, useful in grades 3 through 8 in the elementary school. The Unit Scales of Attainment and the New Stanford Achievement Test include tests of

¹ Devised by S. L. Pressey, L. C. Pressey, Helen Ruhlen, E. V. Bowers, Blythe Pearce, and F. R. Conkling and published by the Public School Publishing Company, Bloomington, Ill.

² Devised by J. P. Leonard and published by the World Book Company, Yonkers-on-Hudson, N. Y.

³ Devised by Caroline Armstrong and published by the California Test Bureau, Los Angeles.

literary information which should prove useful in the upper elementary grades.

*The Wilson Language Error Test.*¹—The purpose of the tests that have been discussed in the preceding paragraphs is the identification of pupil errors in the mechanics of written expression and in grammatical usage. The Wilson Language Error Test also is based on the specific language errors of pupils. The test material is given in the form of a pupil's composition, thus presenting the pupil with the situation that he encounters when he is attempting to correct something that he has written himself. Essentially, the test is a proofreading test. The form of the instrument is best shown by presenting one of the stories from the test.

STORY D

Playing Marbles

Marbles is a good game. I seen some boys playing the game yesterday. I went home to look for my supply of marbles. I couldn't find none, so I went to my father. I said to him, "Father, I ain't got no marbles. Will you give me a dime?" Father saw that I was in earnest, so he give me a dime. He done it willingly. Me and father is very good friends.

I started down the street. I had not went very far, when I met John Taylor. John he is a good friend of mine. He had saw me leave are house, and had came to meet me. I owed him a dime but he did not ask me to pay him. I suppose he wished me to have some marbles so as I could play with him. He had some marbles hisself. He is awful good to me.

I ask him to go to the store with me. "I can't," he said, "I have got an errand to do. Can I play with you when I have did the errand?" We agreed and spent the hole afternoon together. We had a good time. We expect to play agin. If I was him, I would practice more. Their is no other way to become a good player. I git some practice every day.

This test provides an analysis of the ability of pupils to recognize common errors in written expression. It provides accurate information as to what should be taught to pupils, in attempting to make their knowledge of mechanics and of grammar function properly in written expression. The analysis is important, since it precedes systematic drill in eradicating weaknesses. Norms are available for grades 3 through 8 in the elementary school.

¹ Devised by G. M. Wilson, Alice Dunn, and Eleanor J. Murphy and published by the World Book Company, Yonkers-on-Hudson, N. Y.

The purpose of testing language abilities is to locate errors, to obtain specific information as to the type and number of errors made as a basis for remedial instruction. The teacher who uses a survey test will discover whether or not her class is up to par in the abilities measured and whether or not diagnostic testing and remedial teaching are essential. Systematic measurement is the basis of good teaching, and, indeed, it guides the direction and the amount of instruction in specific subject matter.

Readings for Further Study

1. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. IX.
2. W. S. MONROE, J. C. DEVOSS, and F. J. KELLY, *Educational Tests and Measurements* (rev. ed.). Chap. VI.
3. E. M. PAULU, *Diagnostic Testing and Remedial Teaching*. Chap. VIII.
4. S. L. PRESSEY and L. C. PRESSEY, *Introduction to the Use of Standardized Tests* (rev. ed.). Chap. VII. Yonkers-on-Hudson, N. Y.: World Book Company, 1931.
5. H. L. SMITH and W. W. WRIGHT, *Tests and Measurements*. Chap. VII. New York: Silver, Burdett & Company, 1928.
6. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chaps. XII, XIII.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. VI, VII.

IV. READING

The preceding discussion has dealt entirely with aspects of written expression, which is essential in transmitting ideas from one individual to another. The discussion that follows is concerned with the measurement of reading. It is apropos first to consider briefly the significance and the importance of reading.

Reading is of great importance to every individual, not so much in itself but because it is the vehicle for obtaining ideas that cannot be transmitted verbally. In so far as schooling is concerned, reading is the fundamental and most important study tool. Academic training places great emphasis on knowledge, most of which comes to pupils from textbooks and other printed sources. The ability to read and to comprehend the meaning of printed English language is therefore a prerequisite to academic success. In much the same manner, reading influences vocational success in adult living, and, also, it plays its significant part in adult recreation.

Because of its usefulness to all individuals throughout their life, it is highly important that reading shall be well taught in the

schools. This is true particularly because in the great majority of school systems reading is taught only in the elementary schools. This means that in educational organizations including junior high schools, pupils progressing at the normal rate will receive only six years of definitely planned instruction in reading. Adequate testing ordinarily means good teaching, so systematic testing is essential to successful instruction in reading. Probably good testing is more important in reading than in any other subject in the elementary-school curriculum.

The Elementary-school Program of Teaching and Testing in Reading.—Grade 1 is the initial reading period, largely devoted to training in reading readiness during the first months but progressing to book reading before the end of the school year. Grades 2, 3, and 4 provide the opportunity for initiating and developing the fundamental attitudes, habits, and skills. During this time, the pupil develops his vocabulary, masters and perfects the mechanics of eye movements, and develops proper pronunciation and enunciation, becoming fluent in both oral and silent reading. The upper grades, 5 through 8, continue the work done in the lower grades, particularly for retarded readers, but these should be thought of normally as the period of extensive reading and for the refinement of reading technique. The majority of reading done in the primary grades necessarily is oral reading, and silent reading is emphasized similarly in the upper grades. The transition from the predominance of oral reading to the predominance of silent reading occurs during the middle elementary grades.

The program of measurement in reading in any given school system depends upon many factors: the characteristics of the pupils, the skill of teachers, the financial resources of the school system, the size of classes, adequate equipment for teaching, the availability of specialists for clinical work of various types, and the like. No attempt will be made to do more than outline a minimum program of measurement for a typical school system.

The program of measurement in reading when pupils enter grade 1 should include a test of hearing¹ and tests of vision, mental ability, and readiness to read. Measurement of mental ability will be discussed in Chapter XII. The testing of hearing

¹ The Maico D-6 Audiometer, manufactured and distributed by The Maico Company, Inc., Minneapolis, Minn., is an easily used instrument for measuring pupils' hearing.

and of vision should be done by a school medical staff. The telebinocular tests for visual sensation and perception will be described later in this chapter. Reading readiness tests also will be discussed later in this chapter.

Tests of comprehension, particularly word recognition tests, and tests of oral reading ability should be given at the end of the work in grade 1. In grades 2 and 3, silent reading tests should be given which include exercises in word meaning, sentence meaning, and simple paragraph reading. The sentence meaning and paragraph reading tests should include exercises requiring the following of precise directions and those requiring the comprehension of specific facts. The use of oral reading tests is important in these grades; indeed, these instruments should be used in all elementary grades through 8. In grades 3 and 4, more difficult materials should be included in the test content, together with exercises requiring comprehension of central thought and of total paragraph meaning. In these grades, also, attention should be given to measuring speed of reading, and this should be continued through the upper elementary grades.

In grades 5 through 8, the types of measurement utilized in the lower elementary grades should be continued, on a difficulty level adequate to the abilities of pupils in the several grades. The ability to organize materials read should be added to those previously mentioned specifically. Attention should be given, apart from tests of comprehension of the printed English language, to measurement of specialized skills, such as ability to use the dictionary and both general and specialized reference source materials. The tests used in these grades should indicate speed of reading and the accuracy and difficulty level of pupils' comprehension during silent reading. Tests should be used in measuring pupils' general competence and specific limitations in oral reading, but this phase of measurement is definitely less important in the upper grades than is the measurement of silent reading, since it is upon silent reading that emphasis is placed in these grades.

The following pages present information concerning typical specific instruments that may be used in carrying out such a program of measurement. There is a wealth of excellent material for use in testing achievement in reading in the elementary grades, and it is not possible to discuss more than a very few

instruments which are available. The discussion is intended to indicate the nature of reading tests in general rather than to recommend the use of any one or more of the measuring instruments described. There are alternate similar tests which are as valuable as those which have been treated in the following discussion.¹

The discussion will be organized in four divisions including, in order, tests of reading readiness, performance tests of fundamental skills necessary in reading, and achievement tests (1) in primary grades and (2) in upper elementary grades.

Reading Readiness Tests.—These tests attempt to predict the success that pupils will have in learning to read. In general, they attempt to indicate which pupils will succeed in learning to read, those for whom success is doubtful, and those for whom failure is certain at the given time. There is little effort to attempt point-for-point prediction of success, but rather the effort has been to establish critical score levels, which serve as barriers to divide pupils of differing abilities from each other.

*The Lee-Clark Reading Readiness Test.*²—This instrument includes four subtests, of which the following materials are samples:

	Test 1				Test 2	
	X		S		s	g
	S		G		m	u
	Test 3				Test 4	
X	X	R	X	am	amd	
S	A	S	S	fun	futn	

Tests 1 and 3 utilize capital letters, of two sizes in Test 3; Tests 2 and 4, lower case letters. In Tests 1 and 2, the pupil is to draw lines connecting the letters that are alike; in Tests 3 and 4, to cross out the letter that does not belong with the others.

This test does predict reading achievement very satisfactorily. It is a group test which may be given in a minimum time, 15 minutes.

Other Tests of Reading Readiness.—The Van Wagenen Reading Readiness Test³ is made up of six subtests, each measuring a

¹ This statement applies obviously to the discussions of tests in this and following chapters, but it is particularly apropos in the case of tests in the fields of reading and arithmetic because of the many that are available.

² Devised by J. M. Lee and W. W. Clark and published by the California Test Bureau, Los Angeles.

³ Devised by M. J. Van Wagenen and published by the Educational Test Bureau, Inc., Minneapolis, Minn.

different specific function: the pupil's range of general information, his ability to detect and to use relationships, his oral vocabulary, his memory span for ideas, his ability to discriminate between words, and the number of repetitions that he needs in order to learn the names of printed words. This is an excellent test of reading readiness, and its value is enhanced by the existence of two forms and of a profile chart. Unfortunately, it is an individual test, requiring considerable time in the testing of large class groups.

The Metropolitan Reading Readiness Test¹ includes seven subtests: (1) perception of similarities of pictures, pictorial symbols, numbers, letters, and words; (2) perception of and copying of designs; (3) oral vocabulary, with word recognition indicated by the marking of pictures; (4) following directions, given in sentences dictated orally, with understanding indicated by the marking of pictures; (5) a test of numbers and number concepts;

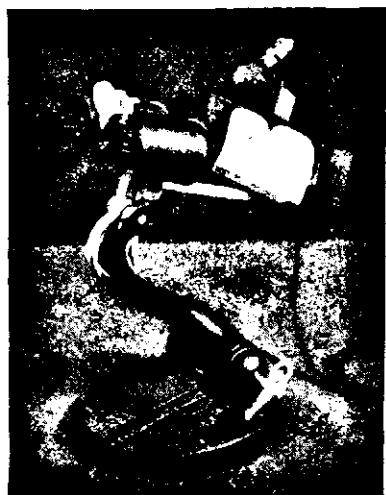


FIG. 7.—The Betts-Keystone telebinocular.

(6) general information; and (7) the drawing of a man. This also predicts readiness very well. It is a group test, requiring 70 minutes approximately in administration, and the work should be done at several different times.

The first of the performance tests described below also yields an index of readiness to read.

Performance Tests. *The Betts Ready to Read Tests.*²—The materials in this battery include tests to be used in diagnosing causes underlying the difficulties of pupils who fail to read normally, and tests

of reading readiness. In addition to tests of auditory readiness and of handedness, the battery includes tests of visual readiness,

¹ Devised by Gertrude H. Hildreth and Nellie L. Griffiths and published by the World Book Company, Yonkers-on-Hudson, N. Y.

² Devised by E. A. Betts and distributed by the Keystone View Company, Meadville, Pa.

of visual sensation and perception, and of oculomotor and perception habits. The last two consist of a series of stereoscopic slides which in their administration require the use of the Betts-Keystone telebinocular, a modification of the familiar stereoscope.

The visual readiness tests measure a child's ability to discriminate visually among readily confused or reversed letters, phonetic elements, and words. The series includes two tests of letter forms and one test each of word forms, phonetic elements, and letter recognition. For this series, the telebinocular is not necessary.

The visual sensation and perception series includes eight tests, of which the first is useful only in establishing the existence of

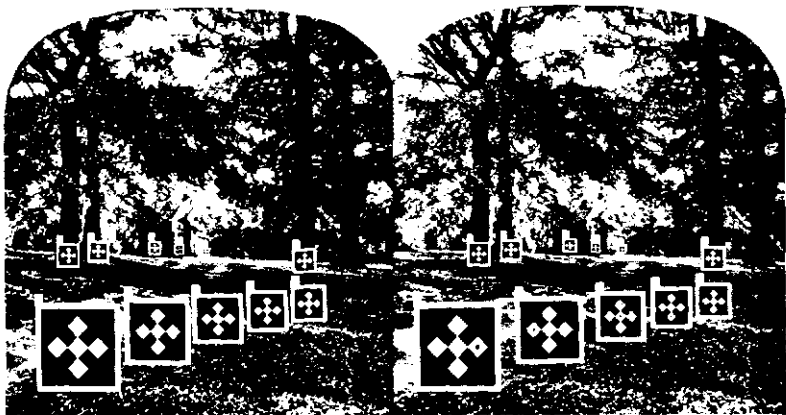


FIG. 8.—Slide DB-3, visual efficiency, right eye, of the Betts Visual Sensation and Perception Test.

binocular vision. The remaining seven tests require the use of 10 slides and, if the color-vision test is given, two additional slides. These tests measure near-point (reading distance) and far-point fusion; visual acuity of the right, left, and both eyes; vertical and lateral imbalance of the eye muscles; eye-coordination level, the degree of stereopsis or depth perception; and sharpness of image, indicating the existence or absence of astigmatism in each eye. This series of tests is more extensive than the usual eye inspection, in which ordinarily only the Snellen Chart is used, measuring visual acuity. These tests do not define and diagnose visual defects with sufficient exactness for prescription; rather, they identify probable defects and indicate

those pupils who should be referred to the eye specialist for examination and treatment, if necessary.

The oculomotor and perception habits series includes eight tests of word orientation, three of gross binocular vision, two of superimposition, and three of stereopsis; one test of letter orientation and one of number orientation, both of gross binocular vision; and a test of ocular dominance. The slides contain many reversible, isolated words, letters, and numbers. Accurate responses demand good eye coordination, fusion, and depth perception.

Norms are available for the visual and auditory readiness tests and for the oculomotor and perception habits tests.

The visual and auditory readiness tests are primarily tests of maturation. They indicate whether or not a child is ready to read. The visual sensation and perception tests are used in detecting functional difficulties of seeing. The oculomotor and perception habits tests are unique word perception tests which determine a specific difficulty and are used for securing an index of general level of achievement when a pupil has less than normal fourth-grade reading ability. Some authorities have questioned the value of certain of these tests; but when they are properly used, they yield valuable information. Probably too little information is now available concerning the validity and the reliability of these instruments. Clinical workers would welcome more statistical data and information concerning the standardization of the tests and their predictive values.

*The Ophthalmograph.*¹—This instrument is a portable, binocular eye-movement camera. It is possible to use it in obtaining quite simply a photographic record of the eye movements of both eyes during binocular reading. The binocular reading graph, or eye-movement photograph, provides objective information concerning the number of fixations per line of type, the regressive movements, the speed of reading, and the coordination of the two eyes during the reading of printed English subject matter. This information is necessary for an evaluation of an individual's efficiency in the mechanics of reading. The reading graph supplements the ordinary eye examination, since it shows the behavior of the eyes while they are at work.

Obviously, the data yielded by the ophthalmograph indicates reading status rather than proficiency in reading. Comprehen-

¹ Distributed by The American Optical Company, Southbridge, Mass.

sion is not measured adequately, and the binocular reading graph data must be supplemented by comprehension measures yielded by standardized tests. Nevertheless, the ophthalmograph has definite clinical values as a prognostic and diagnostic instrument.

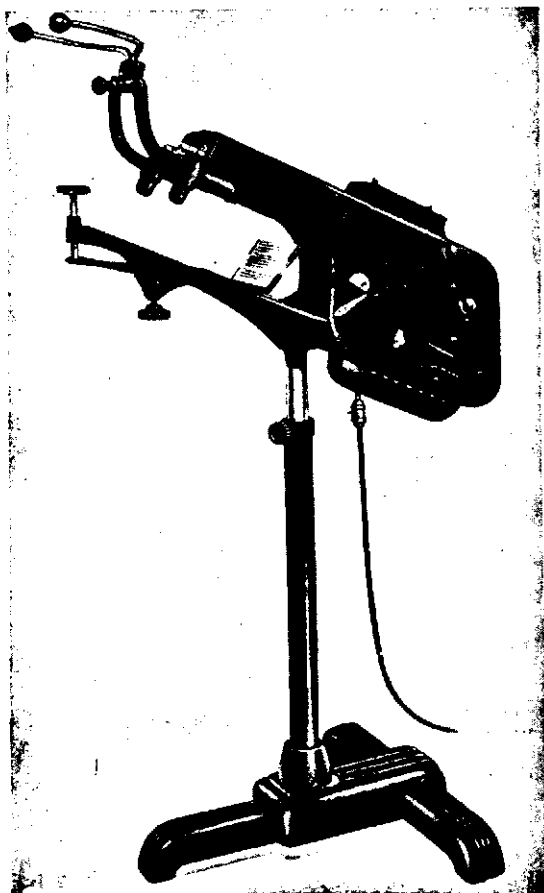


FIG. 9.—The Ophthalmograph.

Although a majority of the authorities have testified as to the reliability and validity of the binocular reading graph, there is some contradictory evidence bearing on the ophthalmograph procedure as it is now used.¹ Apparently, changing the pro-

¹ H. A. IMUS, J. W. ROTHNEY, and R. M. BEAR, *An Evaluation of Visual Factors in Reading*, Hanover, N. H.: Dartmouth College Publications, 1938.

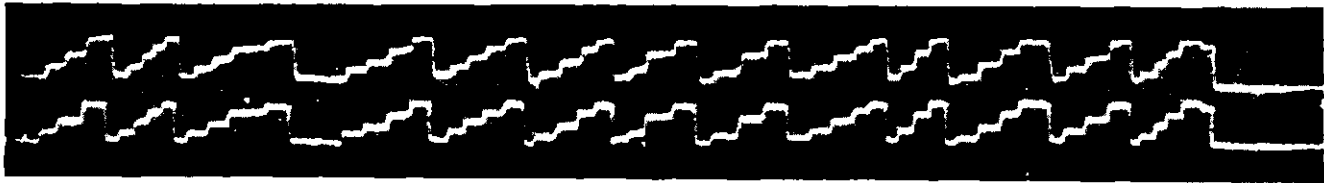


FIG. 10. The binocular reading graph of an efficient reader. Only about 25 per cent of persons tested have acquired, without special training, the rhythmical, coordinated eye-movement habits revealed in this reading graph.

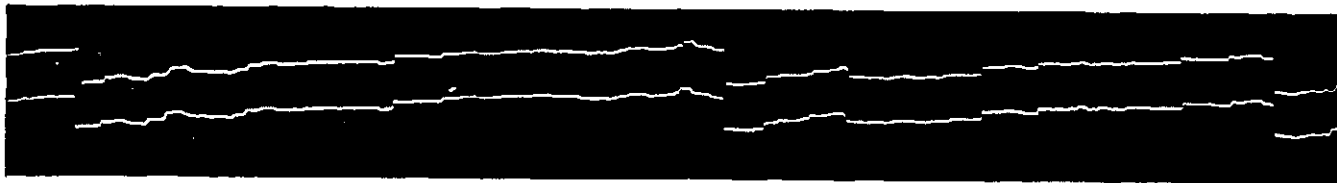


FIG. 11.—The binocular reading graph of an inefficient reader. This pupil was in grade 5 after failing three times in covering five years of elementary school work. The graph reveals: (1) a very narrow span of recognition, indicated by the excessive number of fixations; (2) about 300 fixations and 125 regressions per 100 words, when the grade 5 norm is 125 fixations and 25 regressions per 100 words; and (3) a speed of about 20 words per minute. The grade 5 norm is 190 words per minute. In terms of fixations, regressions, and reading rate, this performance is below the average for a grade 1 pupil who has finished one semester of reading training. (These graphs and data furnished by E. A. Taylor, American Optical Company, Southbridge, Mass.)

cedure, particularly by lengthening the passages that are read, would remove much of the apparent causes of low reliability and validity.

The remaining two divisions are concerned with achievement tests primarily, those of which classroom teachers ordinarily think when they think of the measurement of reading. The discussion will be concerned first with the measurement of achievement in reading in primary grades, and this will be followed by a discussion of the measurement of achievement in reading in grades above the primary. Both oral and silent reading tests will be discussed.

The Measurement of Oral Reading. *The Gray Standardized Oral Reading Check Tests.*¹—This is a series of tests, including printed materials at four levels of difficulty. Set I is intended for use in the first grade; Set II, in the second and third grades; Set III, in the fourth and fifth grades; and Set IV, in the sixth, seventh, and eighth grades.² Each set contains five tests of about equal difficulty. Those in each set are numbered 1 to 5 inclusive. Each test in Set I contains 40 words printed in nine short lines. Each test in the remaining three sets contains three paragraphs of 50 words each.

The purposes of this test series are (1) to secure accurate measures at frequent intervals of the progress of pupils in rate and accuracy of oral reading and (2) to secure detailed information which will aid in determining specifically the nature of the difficulties of poor oral readers.

The character of the test content may be illustrated by presenting Test 1 of Set II.

SET II—No. 1

A nest is in a big green tree. The mother bird made the nest. She put it on the branch of the tree among the pretty leaves. She made it of twigs, leaves, and grass. She put

¹ Devised by W. S. Gray and published by the Public School Publishing Company, Bloomington, Ill.

² It should be noted that this test series is introduced under the primary grades, although it includes upper grade materials, because (1) oral reading is stressed in primary grades, and (2) much of the test material is intended for use in these grades.

soft rags inside of it. The nest has five baby birds in it.

The nest is large and round. The little birds will not fall out. The nest holds the mother bird and the little birds, too. It is hidden under the leaves. The old cat cannot see it. He does not know where the birds are. He will not find them there.

The nest is the home of the birds. It is a bed for the baby birds. The wind rocks it back and forth. The nest is very strong and the wind cannot blow it down. The little birds eat and sleep all day. They will learn to fly very soon.

The tests must be given to pupils individually, since it is necessary for the examiner to record the time required by the pupil in reading the selection, together with the errors made during the reading. The types of errors and the manner of recording them are clearly indicated in the following paragraphs, which are taken from the directions accompanying the test series.

The sun pierced into my ^{many} large windows. It was the opening of October, and the ^{clear} sky was of a dāzzling blue. I looked out of my window and down the street. The white houses of the long, straight street were almost painful to the eyes. The clear atmosphere allowed full play to the sun's brightness.

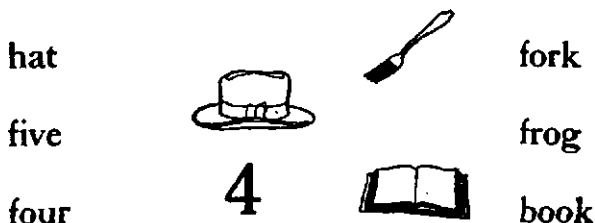
If a word is wholly mispronounced, underline it as in the case of "atmosphere." If a portion of a word is mispronounced, mark appropriately as indicated above, for example, "pierced" pronounced in two syllables; sounding long a in "dazzling;" omitting the s in "houses," the al in "almost," or the r in "straight." Omitted words are marked as in the case of "of" and "and;" substitutions as in the case of "many" for "my;" insertions as in the case of "clear;" and repetitions as in the case of "to the sun's." Two or more words should be repeated to count as a repetition.

Record the exact nature of each error as nearly as you can. When you are unable to define clearly the specific character of an error, underline the word or portion of the word mispronounced. Be sure you put down a mark for each error. In case you are not sure that an error was made, give the pupil the benefit of the doubt. If the pupil has a slight foreign accent, distinguish carefully between this difficulty and real errors.

Midyear norms are available, both for speed and for errors in all grades.

This series of tests is valuable in guiding remedial reading instruction of poor oral readers. The identification of specific errors is a necessary first step in the correction of such errors.

The Measurement of Silent Reading in Primary Grades. *The Los Angeles Primary Word Recognition Test.*¹—This test is typical of the numerous word recognition tests that are available. It purports to yield a measure of the ability of pupils to identify printed English words. Essentially, it is a vocabulary test, although only in part does it require a knowledge of word meaning, in the matching of words and pictures. The test contains two parts: (1) a matching exercise, in which the pupils draw lines



from the word to the related picture, and (2) and a dictation test, in which the pupils identify the word dictated from among four words given. The excerpts from the test indicate the character of the materials

ball	bird	big	blue
goat	good	go	girl

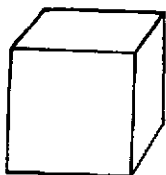
*The DeVault Primary Reading Test.*²—This test contains three subtests. Each of these requires some sentence reading. In addition, the first division deals also with the identification of certain words and with following directions which are dictated orally. The second division is a general information test in sentence form. The third presents sentence materials, measuring the following of directions, and paragraph materials dealing with

¹ Devised by the Research Division, Los Angeles Public Schools, and published by the California Test Bureau, Los Angeles.

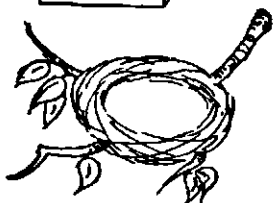
² Devised by Nellie M. DeVault and published by the California Test Bureau, Los Angeles.

specific facts. The following excerpt is from the first items in the third division.

1. Make a ball in the box.



2. Make one egg in the nest.



This is an excellent primary reading test. When two forms are used, an error analysis of the tests provides a diagnosis adequate for the guidance of remedial instruction in silent reading.

*The Williams Primary Reading Test.*¹—Like the DeVault Primary Reading Test, this instrument deals with sentence and paragraph reading. The sentence division deals with following directions; and the paragraph materials, principally with specific facts. The paragraph-reading portion of the test is much longer than the sentence-reading portion. The paragraphs describe situations common to most first- and second-grade readers, except the last and most difficult paragraphs. These are so difficult that but few third-grade pupils can complete the entire test successfully.

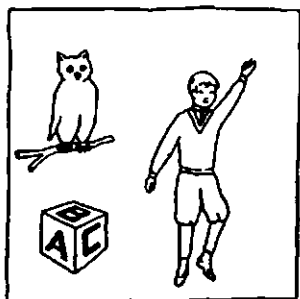
The types of content of this test are illustrated at the top of page 196.

*The Gates Silent Reading Tests: Primary Series.*²—This series of tests includes three instruments measuring separately (1) word recognition, (2) sentence reading, and (3) paragraph reading. The tests may be used separately, although they were designed for use as a team. When used together, they should be administered preferably in the order in which they are numbered, and the total composite score yields a general measure of reading

¹ Devised by A. J. Williams and published by the Public School Publishing Company, Bloomington, Ill.

² Devised by A. I. Gates and published by the Bureau of Publications, Teachers College, Columbia University, New York.

1. Put a cross on the owl.
2. Make a line under the boy.
3. Put a ring around the block.



Read this story and answer the questions about it.

A boy found a rabbit in the road. He took it home with him and gave it some clover to eat.

4. What did the boy find?.....
5. Where did he take the rabbit?.....

achievement. The tests attempt to secure an objective diagnosis of strengths and weaknesses in the most significant reading abilities in the primary grades. These instruments are among the better primary reading tests.

A Comparison of Primary Reading Tests.—It is obvious from the brief preceding discussions that primary reading tests measure similar things to a marked extent but that not all such instruments measure all types of achievement in primary reading. Furthermore, there are differences in the difficulty range of different tests. Data normally are available as to grade placement equivalents to raw scores and reading age. Some evidence is available to show that norms of certain of the primary reading test are comparable.¹ Evidence available indicates that measures yielded by primary reading tests show a high degree of relationship; *i.e.*, the tests validate each other quite well. Typical findings are given in Table XVIII. The tests used in the study from which Table XVIII was taken were the Los Angeles Primary Word Recognition Test, Form I; the Detroit Reading Test, Test I, Form A; the DeVault Primary Reading Test, Form 1; the Los Angeles Primary Reading Test, Form I; the Williams Primary Reading Test, Form A; and the New Stanford Achievement Test, Primary Examination, Reading Section, Form V.

¹ *E.g.*, see NELLIE M. DEVAULT, *Manual of Directions for the DeVault Primary Reading Test*, p. 3, Table III.

TABLE XVIII.—MEASURES OF RELATIONSHIP BETWEEN SCORES YIELDED BY SIX PRIMARY READING TESTS¹

N = 87

	Los Angeles Primary Reading Test	Detroit	DeVault	Williams	New Stanford
Los Angeles Primary Word Recognition Test.....	0.665	0.728	0.824	0.716	0.657
Los Angeles PRT.....	0.631	0.633	0.751	0.677
Detroit.....	0.578	0.755	0.571
DeVault.....	0.781	0.599
Williams.....	0.537

¹ F. BAKER and M. E. BROOM, Concerning One Criterion for the Choice of Primary Reading Tests, *Journal of Applied Psychology*, 16: 419-420 (August, 1932).

The data in Table XVIII indicate that these six instruments validate each other quite well and that whatever is measured by any one of them is measured to a marked degree by each of the others. Similar findings have been reported by other investigators.¹ This does not mean that one may use any one of the many primary reading tests that are available and secure valuable data in a specific problem situation; it means rather that, in a normal situation, one may use any one of several similar instruments and be assured of obtaining a valid and reliable measure of reading achievement in primary grades. The selection of standardized tests should always be made with all factors taken into account, in any case, and this means a careful study of both the situation and the tests.

The Measurement of Silent Reading in the Intermediate and Upper Grades. *The Los Angeles Elementary Reading Test.*²—

¹ *E.g.*, see ELLEN McANULTY and W. W. CLARK, A Study of Primary Reading Tests, *Los Angeles Educational Research Bulletin*, 10 (No. 5): 6-16 (January, 1931).

LOUISE E. WELLER, The Improvement of Silent Reading, pp. 87, 89, 90, 92, 94, 95, unpublished Master's thesis, University of Southern California, 1932.

² Devised by Jessie E. Ingraham and published by the California Test Bureau, Los Angeles.

This is a group test, intended for use in measuring silent reading comprehension in grades 3 through 8. The test materials are scaled in order of difficulty. The test items call for a variety of reading comprehension skills, including the knowledge of word meanings, the recognition of directly stated facts, the drawing of inferences, the discrimination between true and false facts, the ability to follow printed directions, and the ability to perceive relationships.

A typical section from the test content follows.

Steve stood beside his cherished cat-boat, "The Sandpiper," as she lay drawn up on the beach, and looked indignantly at her scratched paint and the deep dent in her side.

"That's what happens when babies try to be sports:" he cried angrily.

14. Steve owned a _____

Draw a line under the words below that tell the truth :

15. Steve valued his boat

just a little
very highly
some what

16. Steve was angry

because bad boys had spoiled the paint

because someone had pounded his boat

because some children had tried to sail his boat

In the manual accompanying the test, a tabulation is given of the abilities measured by the test.¹ This is useful in directing the teacher's attention to the large number of abilities that are commonly grouped together when one thinks of reading and to the specific character of these abilities. The analysis is not particularly valuable for use in diagnosing pupils' reading difficulties, because of the limited number of items under several of the different types. For example, there is only one item that demands the ability to perceive an analogy.

This test is an excellent survey instrument, but it does not yield enough information for diagnostic purposes.

*The Sangren-Woody Reading Test.*²—This instrument consists of a series of seven parts so constructed as to measure seven

¹ JESSIE E. INGRAHAM, *Manual of Directions for the Los Angeles Elementary Reading Test*, p. 4.

² Devised by P. V. Sangren and C. Woody and published by the World Book Company, Yonkers-on-Hudson, N. Y.

distinct phases of silent reading ability. The seven different phases of silent reading measured separately in the test are (1) word meaning, (2) rate or speed of reading, (3) comprehension of specific facts, (4) total meaning, (5) central thought, (6) following directions, and (7) logical organization of ideas. The separate divisions of the test yield specific measures in several important and clearly distinguished phases of silent reading, making possible a comparison of a pupil's achievement in these different types of silent reading in which he must engage during study.

A portion of the word-meaning division of the test appears below. It should be noted, perhaps, that this type of test has not been illustrated heretofore.

11. Almost means very 1 long, 2 high, 3 nearly, 4 pretty.....()
12. To settle is to 1 buy, 2 please, 3 enjoy, 4 stop.....()
13. To treat is to 1 run, 2 hurt, 3 play, 4 entertain.....()
14. The breast of an animal is the 1 back, 2 head, 3 foot, 4 front..()
15. To wreck is to 1 ruin, 2 occur, 3 trust, 4 refuse.....()

The use of a single form of this test does not yield information in sufficient quantity to permit a reliable diagnosis; for this purpose, two forms should be used. Nevertheless, the test does indicate quite well the strengths and weaknesses of groups in the several phases of silent reading when only one form is used. This serves as a basis for planning remedial instruction methods and materials.

*The Gates Silent Reading Tests: Series for Grades 3 through 8.*¹—This series of tests, intended for use in grades 3 through 8, includes four instruments: (1) a test of reading to appreciate general significance, (2) a test of reading to predict the outcomes of given events, (3) a test of reading to understand precise directions, and (4) a test of reading to note details. These four tests provide an accurate measure of reading ability in general, and they yield also a diagnosis of particular strengths and weaknesses in certain reading skills. A knowledge of the types of reading ability in need of development permits the planning of instruction to improve specifically those types of technique. Since each test in the series measures a particular type of reading skill that

¹ Devised by A. I. Gates and published by the Bureau of Publications, Teachers College, Columbia University, New York.

is important, and a type that can be acquired, this team of tests is of value to classroom teachers of reading.

*The Stone Narrative Reading Tests.*¹—The majority of reading tests for use in intermediate and upper elementary grades are paragraph tests. These narrative reading tests measure ability to read complete narratives. Three types of comprehension are measured corresponding to three essential elements of a good narrative: (1) comprehension of incidents essential in developing a real point or plot, (2) comprehension of incidents essential in developing steps in the plot or story, and (3) comprehension of incidents that fill in and enrich the narrative with details.

Separate tests are available for each grade of the elementary school in grades 4 through 8. There is a preparatory test for each grade, shorter than the real test but similar otherwise. There is a comprehension test of 20 questions for each selection. Each question has been typed under the three types of comprehension mentioned in the preceding paragraph. Two diagnostic charts aid the teacher in studying pupils' responses to the test.

These tests are concerned with a specific type of reading material, and the results indicate a specific type of comprehension ability. This test supplements the types of tests discussed previously.

A Comparison of Intermediate- and Upper-grade Reading Tests.—This discussion excludes the Stone Narrative Reading Tests, because of their difference from others in type and extent of content. The intermediate- and upper-grade reading tests, measuring comprehension of the printed English language, include similar types of content and similar exercises to a large extent. There are, of course, differences in the difficulty of the materials of different tests, and some tests include measures of fewer or of more types of reading ability than do certain others. Derived scores and norms commonly are available. There is some evidence that grade placement scores yielded by different reading tests are not comparable.² Evidence is available proving that measures yielded by tests devised for use in intermediate and upper elementary grades show a high degree of relationship. For

¹ Devised by C. R. Stone, Annette Buehrmann, and Lucile C. Murphy and published by the Public School Publishing Company, Bloomington, Ill.

² JESSIE E. INGRAHAM, *Manual of Directions for the Los Angeles Elementary Reading Test*, p. 3, Table II.

examples of measures of such relationship, refer to Table IX (page 89) and to Table X (page 91), considering standardized test data only.¹

The user of tests measuring comprehension during silent reading in the intermediate and upper elementary grades is faced with the necessity of considering the individuals who are to be tested, the purpose of the testing, and the numerous tests available. Only the careful selection of suitable measuring instruments, adapted to the specific purposes in view, and the precise use of the materials will yield results of any significance or value.

Diagnostic Tests.—There are in existence certain types of diagnostic tests that have not yet been discussed. Essentially, these are individual tests, devised to permit the detection of reading difficulties and their diagnosis in individual cases. It is impossible for anyone to grasp the nature of the testing of reading without at least a brief discussion of these types of tests.

*The Durrell Analysis of Reading Difficulty Tests.*²—This is a battery of individual tests, including oral reading and recall, silent reading and recall, word analysis as measured by tachistoscopic exposures compared with long-time exposures, speed of reading, phonetic abilities, handwriting, and spelling. A check list of difficulties is provided; norms at three achievement levels in each grade, low, medium, and high; and a profile chart. These aids make easy the task of noting errors, mannerisms, and difficulties during reading.

*The Gates Reading Diagnosis Tests.*³—This test is intended for use with other materials in making an extensive diagnosis of the difficulties of handicapped readers.⁴ The reading diagnosis tests include the following: (1) an oral context-reading test; (2) a graded word pronunciation test; (3) tests of word recognition techniques, of ability to recognize a word seen, and of ability to

¹ See also ELLEN McANULTY and W. W. CLARK, *A Study of Elementary Reading Tests*, *Los Angeles Educational Research Bulletin* 10 (No. 6): 2-10 (February, 1931).

² Devised by D. D. Durrell and published by the World Book Company, Yonkers-on-Hudson, N. Y.

³ Devised by A. I. Gates and published by the Bureau of Publications, Teachers College, Columbia University, New York.

⁴ Directions for administering and scoring the tests are available in A. I. Gates, *The Improvement of Reading* (rev. ed.), Appendix I, pp. 503ff., New York: The Macmillan Company, 1935.

recognize a word spoken; (4) tests of phonogram combinations, of initial vowel syllables or initial consonant syllables or vowel-consonant phonograms or vowel phonograms, of consonant phonograms or consonant-vowel phonograms, of blending letter sounds or letter sounds or capital letters, and of lower case letters; and (5) tests of visual perception, same-different figures, same-different numbers, and selection of figures. This series of tests should be of great value in clinical diagnosis of the difficulties of handicapped readers.

Summary.—This division of the discussion of measurement in the language arts has been concerned with reading, the one most important subject of instruction in the elementary-school grades. Tests have been described for use in inspecting visual efficiency, in examining the efficiency of eye movements during reading, in investigating reading readiness, and in measuring oral and silent reading achievement in all elementary-school grades. Certain diagnostic tests have been discussed briefly. The reader may feel that more information should have been given concerning these instruments, but space limitations prevent extended discussions; and in any case, the student should examine completely the tests and test manuals accompanying them. Only by giving careful thought and extended investigation of numbers of good reading tests can a student acquire a sufficient working knowledge of such instruments.

Little attention has been given directly to the uses of reading tests in the preceding discussion. In general, tests are of value in determining the normalcy of pupils approaching reading; in evaluating the difficulties of pupils while they are reading orally and silently, preliminary to remedial instruction; and in obtaining a general measure of reading achievement, particularly during silent reading, by using survey instruments. The literature dealing with instruction in reading is too extensive to attempt its summary here. Perhaps it is sufficient to say that without adequate measurement, effective instruction is difficult if not impossible.

Readings for Further Study

1. E. A. BETTS, *The Prevention and Correction of Reading Difficulties*. Chaps. V, VII, pp. 310-379. Evanston, Ill.: Row, Peterson and Company, 1936.

2. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. VIII.
3. O. K. BUROS *et al.*, *The 1938 Mental Measurements Yearbook*. Pp. 124-127, 132-135. New Brunswick, N. J.: Rutgers University Press, 1938.
4. A. I. GATES, *The Improvement of Reading* (rev. ed.). Chaps. III, V, VI, IX, XI.
5. M. LUCILLE HARRISON, *Reading Readiness*. Chap. IV. Boston: Houghton Mifflin Company, 1936.
6. MAUDE MCBROOM *et al.*, *Newer Practices in Reading in the Elementary School*, 17th Yearbook of the Department of Elementary School Principals, National Education Association, Washington, D. C., 1938. Pp. 248-254, 397-402, 601-607.
7. H. L. SMITH and W. W. WRIGHT, *Tests and Measurements*. Chap. IX.
8. E. A. TAYLOR, *Controlled Reading*. Chaps. III, IV, VII. Chicago: University of Chicago Press, 1937.
9. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. VIII.
10. G. M. WHIPPLE *et al.*, *The Teaching of Reading: A Second Report*, 36th Yearbook of the National Society for the Study of Education. Part I, Chap. XII. Bloomington, Ill.: Public School Publishing Company, 1937.
11. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. V.

Summary of the Chapter.—The preceding discussion has dealt with standardized tests in the language arts. Language is essential to the communication of ideas. One conveys his ideas to others by speaking or by writing and receives the ideas of others by hearing them speak or by reading what they have written. It is difficult, if not impossible, to measure spoken language objectively, so the content of the chapter has been concerned only with written language. The organization has presented in order handwriting; spelling; composition; mechanics and grammar tests, measuring the abilities and skills necessary in communicating ideas to others; and reading, necessary in obtaining ideas from others.

The effort has been made to give a knowledge of the types of measuring instruments available in classroom teaching rather than a comprehensive knowledge of the many excellent tests that are available. Some attention has been given to the uses of findings yielded by tests. It is obvious that a complete analysis of all of the uses of tests and of the methods of using test data cannot be provided in so brief a discussion. The student is referred to the sources listed at the end of each division of the

chapter for further information on the correct procedures in using data yielded by tests in improving classroom teaching in the language arts.

Questions and Exercises

I. *Handwriting.*

1. Secure samples of handwriting from three persons, and score them, using the Ayres Measuring Scale for Handwriting, Gettysburg Edition. Make a diagnosis of defects in these three samples of handwriting, using the Pressey Chart for Diagnosis of Illegibilities in Handwriting.

2. What conditions might cause the handwriting on the fourth page of a pupil's response to an essay examination to be an unfair sample of his penmanship?

3. How should a classroom teacher proceed in developing an informal handwriting scale including samples of the penmanship of her own pupils?

4. Is it desirable to have pupils diagnose their own difficulties in handwriting? Why?

5. How are samples obtained for use in measuring handwriting quality? Handwriting speed?

6. How does the measurement of handwriting quality differ from the measurement of spelling ability? From the measurement of comprehension during silent reading?

7. Check at least three handwriting scales as to their content, directions for using, norms data, and other characteristics. Prepare a tabulation of these characteristics to identify similarities and differences among the charts.

8. List the measuring instruments that should be available to a handwriting teacher who instructs pupils in grades 2, 5, 7, and 8. Be prepared to defend each choice.

II. *Spelling.*

1. Compare and contrast the Buckingham Revision of the Ayres Spelling Scale and the Iowa Spelling Scale. How was each derived? Which is of greater value to a teacher who is instructing grade 5 in spelling during a period of one school year? Why?

2. What differences may be found among the earlier and the more recent instruments devised to measure achievement in spelling?

3. Why should sentence dictation spelling tests not be timed? What can be done by a teacher using such a test to avoid unnecessary waste of class time?

4. Is a spelling examination on which the median score is a perfect score a good measure of ability to spell? Why?

5. Why does a test of very difficult words fail to give a good measure of spelling ability?

6. Outline the procedure to be followed by a classroom teacher in preparing a sentence dictation spelling test, using words chosen from standardized spelling scales.

7. What types of spelling exercises have been used in tests attempting to measure spelling achievement? Suggest other types of exercises that might be so used?

8. How might a classroom teacher discover the reliability of spelling achievement among her pupils without using the correlation technique?

III. *Mechanics and Grammar.*

1. What difficulties will be encountered when one attempts to secure a general measure of language ability?

2. Select three scales devised for use in judging quality of composition writing, and compare and contrast them. Determine which is most useful to a classroom teacher in grade 7.

3. What phases of language instruction are not now adequately measured by standardized measuring instruments?

4. What makes the measurement of achievement in mechanics and grammar difficult?

5. What is the value of a diagnostic test in mechanics of written expression and grammar? Name several such tests of value in elementary-school grades, indicating the reasons for your choice.

6. Why is the measurement of composition ability less valid and reliable than the measurement of achievement in grammar and in language usage?

7. Is it difficult to diagnose the excellences and the defects in pupils' compositions? Outline a procedure for this purpose, using standardized measuring instruments. Criticize your plan.

8. What types of exercises should be included in language usage tests? Name at least one standardized test in which each of the types listed is to be found.

9. Of what value is a measuring instrument such as the Wilson Language Error Test?

10. Study the language tests in the Unit Scales of Attainment and the Metropolitan Achievement Tests. Compare and contrast these examinations.

11. Which of the available standardized instruments for measuring language abilities include tests of grammar? How much grammar should be included in such an instrument devised for use in elementary grades? Why?

12. List at least five of the outstanding language tests. Prepare a tabulation of the characteristics and content of each of these instruments.

IV. *Reading.*

1. What are the types of reading skills of which a pupil must have a command for study purposes in the upper elementary-school grades? Which of the available standardized tests measure these types of reading skill?

2. In what school grade or grades is vocabulary the most important factor in silent reading comprehension? List the tests that probably are most useful in this grade level.

3. Is rate of reading an important factor to be considered during the measurement of silent reading comprehension? Why?

4. Assume that you have been made responsible for establishing a program of measurement in reading for an elementary-school system, including eight grades in six buildings. The total enrollment is 3,423 in the elementary grades, and the population is homogeneous in its language backgrounds. Indicate the tests that should be used both for normal and for handicapped readers. Tell why each instrument is included in your material.

5. How much time at the beginning of a school term should be spent in discovering pupil reading difficulties and in preparing a chart of each pupil's difficulties? To what extent should oral reading tests be used in such a program? Why?

6. If a classroom teacher in grade 6 wished to compare the general reading achievement of her pupils with that of other pupils in the same grade, what test or tests should she use?

7. What elements of silent reading comprehension are measured least adequately in the Sangren-Woody Reading Test? In the Williams Primary Reading Test? In the Los Angeles Elementary Reading Test?

8. If a classroom teacher in upper elementary grades wished to identify the pupils in her group who were most in need of drill in rapid reading, what test or tests should she use?

9. Select two reading tests that validate each other well, for both primary and upper grades. Compare and contrast the content of the tests, and give an explanation for the validity data on the basis of your findings.

10. What is meant by power in reading? What standardized tests measure this type of reading ability?

11. Why is it necessary to have standardized oral reading tests?

12. What is the place of the telebinocular in a program of measurement of reading achievement? Of the ophthalmograph?

13. Is it necessary to use a specialized silent reading test, such as a test of silent reading comprehension in history, in attempting to analyze the study difficulties of pupils in the social sciences? Why?

14. Prepare a tabulation of 10 silent reading tests, devised for use in one or more grades, 1 through 8, listing reliability and validity data, the types of content included in each test, etc. The tests listed should include at least two measuring instruments devised for use in each of the eight grades.

CHAPTER VIII

STANDARDIZED TESTS IN ARITHMETIC

Arithmetic is studied in elementary schools because of its usefulness in certain of life's essential activities. The aim of arithmetic is to develop a knowledge of quantitative terms and of quantitative relationships, to make known the significance and the function of number and quantity in the activities of daily life, and to develop proficiency in performing computations with facility and accuracy. Instruction in arithmetic can function effectively only when pupils know the necessary vocabulary and quantitative concepts. Arithmetic becomes meaningful when pupils know about the activities and the objects with which it deals, thus making computation a means to an end rather than an end in itself.

The essential aim in arithmetic has been to develop ability in computation and in problem solving, and this still is an important aim. A belief has arisen that arithmetic should do more than develop merely computation and problem-solving skills. Certain additional outcomes should result from instruction in arithmetic—desirable attitudes, ideals, appreciations, powers, and social qualities.

Testing is as important in the teaching of arithmetic as in the teaching of other subjects. Yet measurement in arithmetic has not kept pace with the changing philosophy relating to the teaching of the subject. Practically nothing has been done to develop materials to measure the understanding of the uses of number in reading materials involving quantitative concepts, in interpreting information presented graphically, and in applying number and quantity in concrete situations that occur daily in life activities. Present standardized tests in arithmetic are limited to the measurement of skill and power in computation and in solving reasoning problems. Arithmetic is highly objective; a problem is solved either correctly or incorrectly. The available standardized tests emphasize logical and precise thinking and accuracy in computation.

The discussion that follows will present typical standardized arithmetic tests. Diagnostic tests will be discussed first, followed by a treatment of survey tests, including both computation and reasoning problem tests.

Diagnostic Tests.—These instruments attempt to provide an analysis of pupil weaknesses in arithmetic. Diagnosis of computational difficulties is easier than diagnosis of difficulties in solving reasoning problems. There are only four fundamental operations that can be performed in computation in arithmetic: addition, subtraction, multiplication, and division. Also, but four types of numbers are involved: whole numbers, mixed numbers, common fractions, and decimal fractions. Sixteen combinations must be considered in devising a diagnostic test in computation. The instrument must include several problems of each type at each level of difficulty within the range of ability of the pupils to be examined with the test. The task of constructing an adequate diagnostic test is more difficult than one might anticipate, since there are many different specific skills to be mastered by the pupil, although the exact number of such abilities is not known. For example, some evidence exists to show that $6 + 9 = 15$ demands a different ability from that demanded by $9 + 6 = 15$. Certainly, $6 + 6 + 3 = 15$ demands more ability and different abilities from those demanded by either of the preceding examples.

Less is known about the solving of reasoning problems than about solving computational problems, but even more skills are involved in solving verbal problems than in solving computational problems. The first requisite in solving a reasoning problem is the ability to read and to comprehend the printed problem. The following types of reading ability are required: those needed to comprehend word meanings; to comprehend specific facts, particularly numerical facts; and to organize facts obtained by silent reading into a significant whole. In addition, the pupil attempting to solve a verbal problem must recall arithmetical processes, formulas, and the like that appear to have a bearing on the solving of the problem and must be able to identify and to retain those items which actually are necessary in solving the problem, discarding all irrelevant matter. He must be able to perceive relationships between the subject matter of the problem and the subject matter of arithmetic in general; he must be able to select

from arithmetical theory the materials that are relevant in solving the problem; and he must be able to construct a procedure utilizing correctly these two types of subject matter. This is reasoning forward toward a solution of the problem, utilizing all pertinent subject matter.

After the pupil has (1) read the verbal problem carefully, with adequate understanding of each word; and (2) determined what is given in the problem and what is required for its solution; and (3) selected the processes to be used in solving the problem and planned the procedure for the solution systematically, then he is ready to undertake the computation necessary in obtaining the answer to the problem and the check computations. It should be noted that reasoning problems have their own types and sources of difficulty, plus those which are found in solving computation problems. Obviously, reasoning problems are more difficult for pupils than computation problems. This may explain the fact that diagnostic tests including reasoning problems in arithmetic are fewer and, on the whole, less adequate than are diagnostic tests including computational problems in arithmetic.

*The Buswell-John Diagnostic Chart for Fundamental Processes in Arithmetic.*¹—This test includes materials on the four fundamental processes—addition, subtraction, multiplication, and division. There are 23 problems in addition, 22 each in subtraction and multiplication, and 21 in division. These exercises are arranged according to increasing difficulty in each process. The test is given to pupils individually.

The test attempts to provide information on why a pupil fails and how he made his errors. The method involves observation of the pupil's work, his methods and errors being noted on the Diagnostic Chart which reproduces the problems on the Pupil's Work Sheet; and a diagnosis, using the examples listed in the *Manual of Directions*. Pupils' methods of work are so varied and so complex that efficient teaching requires a systematic and organized scheme for diagnosing them. The materials in the Diagnostic Chart and in the *Manual of Directions* present such a systematic plan for diagnosing the various kinds of habits used by pupils in elementary-school grades in working with the four fundamental processes.

¹ Devised by G. T. Buswell and Lenore John and published by the Public School Publishing Company, Bloomington, Ill.

SUBTRACTION: (Place a check before each habit observed in the pupil's work)

<input type="checkbox"/> s1 Errors in combinations	<input type="checkbox"/> s15 Deducted from minuend when borrowing was not necessary
<input type="checkbox"/> s2 Did not allow for having borrowed	<input type="checkbox"/> s16 Ignored a digit
<input type="checkbox"/> s3 Counting	<input type="checkbox"/> s17 Deducted 2 from minuend after borrowing
<input type="checkbox"/> s4 Errors due to zero in minuend	<input type="checkbox"/> s18 Error due to minuend and subtrahend digits being same
<input type="checkbox"/> s5 Said example backwards	<input type="checkbox"/> s19 Used minuend or subtrahend as remainder
<input type="checkbox"/> s6 Subtracted minuend from subtrahend	<input type="checkbox"/> s20 Reversed digits in remainder
<input type="checkbox"/> s7 Failed to borrow; gave zero as answer	<input type="checkbox"/> s21 Confused process with division or multiplication
<input type="checkbox"/> s8 Added instead of subtracted	<input type="checkbox"/> s22 Skipped one or more decades
<input type="checkbox"/> s9 Error in reading	<input type="checkbox"/> s23 Increased minuend digit after borrowing
<input type="checkbox"/> s10 Used same digit in two columns	<input type="checkbox"/> s24 Based subtraction on multiplication combination
<input type="checkbox"/> s11 Derived unknown from known combination	
<input type="checkbox"/> s12 Omitted a column	
<input type="checkbox"/> s13 Used trial-and-error addition	
<input type="checkbox"/> s14 Split numbers	

Habits not listed above _____

(1)	$\begin{array}{r} 5 \quad 8 \\ \underline{3} \quad \underline{8} \end{array}$	(5)	$\begin{array}{r} 36 \quad 79 \\ \underline{21} \quad \underline{24} \end{array}$
(2)	$7 - 1 =$ $9 - 0 =$	(6)	$\begin{array}{r} 12 \quad 10 \\ \underline{6} \quad \underline{2} \end{array}$
(3)	$\begin{array}{r} 19 \quad 15 \\ \underline{2} \quad \underline{4} \end{array}$	(7)	$\begin{array}{r} 15 \quad 19 \\ \underline{13} \quad \underline{12} \end{array}$
(4)	$\begin{array}{r} 58 \quad 79 \\ \underline{4} \quad \underline{3} \end{array}$	(8)	$59 - 2 =$ $86 - 4 =$

The excerpt from the Diagnostic Chart on page 210 lists 24 types of errors commonly found in pupils' subtraction. The problems are those which appear in the Pupil's Work Sheet for this process. Note that space is available for the examiner to note beside each problem the pupil's work habits and other pertinent data.

This test is valuable in detecting the difficulties of pupils who are not doing satisfactory work in arithmetic. The purpose of this instrument is to give a pupil the greatest possible individual help in improving his work. The attempt is to find out just how the pupil works when he is working independently. The examiner must use the technique of close observation and of skilled questioning as the pupil does his work. After the difficulties are known, the teacher can plan to eradicate poor work habits and incorrect procedures. The procedure with this instrument is intended to obtain essential information prior to teaching.

*The Compass Diagnostic Tests in Arithmetic.*¹—There are 20 tests in this series, each covering from one to seven basic arithmetic processes or skills commonly taught in grades 2 to 8. More than 90 different aspects of arithmetic are measured in the 20 tests.

Each test represents a working analysis of the fundamental unit-abilities making up the larger ability designated by the title of the instrument. In most of the 20 tests, the diagnosis has been carried out to such a fine degree that each subdivision of the test parallels a unit of instruction. This permits the teacher to know of the existence and the specific location of pupils' difficulties.

In order that the student may comprehend the character and scope of these tests, it is necessary to give a catalogue of the 20 tests, part by part. This permits a comprehension of the diagnostic testing that may be done with this series. The grades for which each test was devised and standardized are indicated.

Test I: Addition of whole numbers (grades 2-8)

Part 1: Basic addition facts

Part 2: Higher decade addition

Part 3: Column addition

Part 4: Carrying in column addition

Part 5: Checking answers in addition

Test II: Subtraction of whole numbers (grades 2-8)

Part 1: Basic subtraction facts

¹ Devised by G. M. Ruch, F. B. Knight, H. A. Greene, and J. W. Studebaker and published by Scott, Foresman & Company, Chicago.

Part 2: Harder subtraction

Part 3: Checking answers in subtraction

Part 4: Finding errors in subtraction

Test III: Multiplication of whole numbers (grades 3-8)

Part 1: Basic multiplication facts

Part 2: Additions used in the multiplication of Parts 4, 5, and 6

Part 3: Carrying in addition used in the multiplication of Parts 4, 5, and 6

Part 4: Fundamentals in multiplication

Part 5: Checking multiplication

Part 6: Finding errors in multiplication

Test IV: Division of whole numbers (grades 4-8)

Part 1: The vocabulary of division

Part 2: Fundamentals of short division

Part 3: Short division with carrying

Part 4: Multiplication, addition, and subtraction used in division in Parts 2, 3, 5, 6, and 7

Part 5: Estimating the first quotient figure

Part 6: Fundamentals of long division; checking

Part 7: Finding errors in long division

Test V: Addition of fractions and mixed numbers (grades 5-8)

Part 1: Changing fractions to equivalent forms

Part 2: Finding common denominators

Part 3: Reducing fractions to lowest terms

Part 4: Fundamentals of addition of fractions

Part 5: Reducing answers and finding errors

Test VI: Subtraction of fractions and mixed numbers (grades 5-8)

Part 1: Changing fractions to equivalent forms

Part 2: Finding common denominators

Part 3: Expressing mixed numbers as improper fractions

Part 4: Fundamentals of subtraction of fractions

Part 5: Reducing answers and finding errors

Test VII: Multiplication of fractions and mixed numbers (grades 5-8)

Part 1: Changing mixed numbers to improper fractions

Part 2: Cancellation in multiplication of fractions

Part 3: Reducing answers to best form

Part 4: Fundamentals of multiplication of fractions

Part 5: Finding errors

Test VIII: Division of fractions and mixed numbers (grades 5-8)

Part 1: Changing mixed numbers to improper fractions

Part 2: Cancellation in division of fractions

Part 3: Reducing answers to best form

Part 4: Changing from multiplication to division form

Part 5: Fundamentals of division of fractions

Part 6: Finding errors

Test IX: Addition, subtraction, and multiplication of decimals (grades 5-8)

Part 1: Vocabulary and notation of decimals

- Part 2: Changing fractions and mixed numbers to decimal form, and the reverse
- Part 3: Fundamentals of addition and subtraction of decimals
- Part 4: Place value and pointing in multiplication of decimals
- Part 5: Finding errors in pointing answers
- Text X: Division of decimals (grades 6-8)
 - Part 1: Dividing decimals by pointing
 - Part 2: Locating decimal points in division
 - Part 3: Changing remainders to decimal form
 - Part 4: Fundamentals of division of decimals
 - Part 5: Finding errors
- Test XI: Addition and subtraction of denominate numbers (grades 6-8)
 - Part 1: Knowledge of tables of measure
 - Part 2: Reducing and borrowing in denominate numbers
 - Part 3: Fundamentals of addition of denominate numbers
 - Part 4: Fundamentals of subtraction of denominate numbers
 - Part 5: Finding errors in addition and subtraction of denominate numbers
- Test XII: Multiplication and division of denominate numbers (grades 6-8)
 - Part 1: Knowledge of tables of measure
 - Part 2: Reducing denominate numbers to larger or smaller units
 - Part 3: Fundamentals of multiplication of denominate numbers
 - Part 4: Fundamentals of division of denominate numbers
 - Part 5: Finding errors
- Test XIII: Mensuration (grades 7-8)
 - Part 1: Vocabulary of mensuration
 - Part 2: Mensuration of plain surfaces
 - Part 3: Mensuration of solids
 - Part 4: Finding areas and volumes
 - Part 5: Writing formulas used in mensuration
- Test XIV: The basic facts of percentage (grades 6-8)
 - Part 1: Equivalent fraction, decimal, and percent relations
 - Part 2: Expressing areas in percents
 - Part 3: Choosing correct solutions in percentage
 - Part 4: Easy work in percentage
 - Part 5: Harder work in percentage
- Test XV: Interest and business forms (grades 7-8)
 - Part 1: Vocabulary of interest and business forms
 - Part 2: Common business forms and budgets
 - Part 3: Computation of interest, discount, etc.
 - Part 4: Use of interest tables
- Test XVI: Definitions, rules, and vocabulary of arithmetic (grades 4-8)
 - Part 1: Basic vocabulary and definitions of arithmetic
 - Part 2: Basic rules of arithmetic
- Test XVII: Problem analysis: elementary (grades 5-6)
 - Part 1: Comprehension
 - Part 2: What is given

Part 3: What is called for

Part 4: Probable answer

Part 5: Correct solution

Test XVIII: Problem analysis: advanced (grades 7-8)

Part 1: Comprehension

Part 2: What is given

Part 3: What is called for

Part 4: Probable answer

Part 5: Correct solution

Test XIX: General problem scale: elementary (grades 5-6)

Test XX: General problem scale: advanced (grades 7-8)

These tests meet a number of important criteria. The authors were guided by the following requirements in constructing the tests. These are:

1. It [the test] must cover all of the separate arithmetic skills which have marked social utility and which receive emphasis in the course of study.

2. Each separate skill must be measured accurately enough to insure that the scores will be sufficiently reliable to stand alone. This is equivalent to saying that each part of a test must sample an ability widely enough to furnish an accurate index to the ability of the *individual* pupil.

3. The time required for giving the test must be as brief as is consistent with the requirements already listed. However, the shortest is often not the best one to employ. Adequate sampling of a skill and reliability of results are securable only when sufficient time is devoted to testing.

4. The test must be simple and convenient in administration, and objective in its scoring. The time limits of the various parts must be indicated clearly. The directions to the pupils must be definite. Generally, the procedure should be demonstrated by means of fore-exercises and sample solutions.

5. The test must provide a simple and direct picture of the strengths and weaknesses of the individual pupil.

6. Detailed provision must be made for a *remedial and corrective* program based upon the findings of the diagnostic testing.¹

This series of tests is quite complete, and it will meet any needs for group diagnostic measurement in arithmetic. It is not expected that any school will use all the tests in the series in its testing program. The test authors indicate that the principal

¹ G. M. RUCH *et al.*, *Manual of Directions for the Compass Diagnostic Tests in Arithmetic*, pp. 2, 3.

situations in which these diagnostic tests will be useful are as follows:

1. [When testing] . . . upon completion of some important instructional unit such as long division or addition of fractions in order to determine which pupils are ready to begin a new unit of instruction.

2. [When testing] . . . certain or all of the pupils of a new entering class, particularly when the teacher has no adequate information about the previous preparation of these pupils.

3. [When testing] . . . individual pupils who fail to make normal progress in learning a new skill. Very often the mastery of previously taught topics essential to the new skill has not been secured.

4. [When testing] . . . new pupils for purposes of grade location and classification.

5. [When testing] . . . at intervals entire grades throughout the school system as a check on the general efficiency of arithmetic work in the schools.¹

The fore-exercise of Test XVIII, the advanced problem analysis test, is presented on pages 216 and 217 in order to show the type of test content used and the general arrangement of test items.

The *Manual of Directions* includes general and specific directions for giving the 20 tests and for scoring the tests; norms, both for each test and for the parts of each test; age norms for each total test; data as to diagnostic procedures; and suggestions for a remedial program following diagnosis. This series of tests represents the most comprehensive effort yet made in constructing diagnostic group tests in arithmetic.

Other Diagnostic Tests in Arithmetic.—There are other types of diagnostic arithmetic tests, including much less content. A typical test of this sort is the Upton Inventory Test in Arithmetic.² This instrument, prepared for use in grades 7 and 8, is intended to serve as an inventory of the pupil's knowledge of arithmetic. It covers the essentials of arithmetic for the eight grades of the elementary school but very briefly. The test includes 40 exercises (the final three including five problems each) dealing with fundamental operations with whole numbers, fractions, decimals, and percentages. The 25 concrete or reasoning problems are presented in two parts. It is obvious that so short a

¹ G. M. RUCH *et. al.*, *Manual of Directions for the Compass Diagnostic Tests in Arithmetic*, pp. 12, 13.

² Devised by C. B. Upton and published by the Bureau of Publications, Teachers College, Columbia University, New York.

PROBLEMS

PART 1—COMPREHENSION

PART 2—WHAT IS GIVEN

<p>Read each problem below. Then work across the two facing pages to the right, doing all Parts for one problem before going to the next. Do not go back and work on a Part after you have completed the one following.</p> <p>Read the Sample below.</p>	<p>Put a cross (X) on the line before the one statement below which is true for each problem.</p>	<p>Put a cross (X) on the line before every statement below which tells a fact given in the problem.</p>
<p style="text-align: center;">Sample [Read the problem]</p> <p><i>My reading book has 124 pages. I have read 72 pages. How many pages do I have left to read?</i> →</p>	<p style="text-align: center;">Sample [Check (X) true statement]</p> <p>_____ I have read all my reader. _____ I have read less than half my book. _____ I have the most of my book to read. <input checked="" type="checkbox"/> _____ I have read a little more than half my book. _____ I should add to get the answer to this problem. →</p>	<p style="text-align: center;">Sample [Check (X) what is given]</p> <p>_____ Number of pages to read. <input checked="" type="checkbox"/> _____ Number pages in book. <input checked="" type="checkbox"/> _____ Number of pages I have read. _____ Number of stories I have read. _____ Number of pages with pictures on them. →</p>

PART 3—WHAT IS CALLED FOR

PART 4—PROBABLE ANSWER

PART 5—CORRECT SOLUTION

Put a cross (X) on the line before the one statement below which tells what is called for in the problem.

Put a cross (X) on the line before the one statement below which gives the nearest probable answer to the problem. Do not take time to work the problem.

Put a cross (X) on the line before the one correct solution given for each problem. Figure in the margin if you want to.

Sample

[Check (X) what is called for]

- _____ Number of pages in book.
 Number of pages yet to read.
 _____ Number of pages I have read.
 _____ Number of stories I have read.
 _____ Number of pages with pictures.

→

Sample

[Check probable answer]

- _____ One book.
 _____ About 124 pages.
 _____ About 72 stories.
 About 50 pages.
 _____ About 196 pages.

→

Sample

[Check correct solution]

- _____ $124 + 72 = 196$
 _____ $\left\{ \begin{array}{l} 124 + 72 = 196 \\ 196 \div 2 = 98 \end{array} \right.$
 _____ $\left\{ \begin{array}{l} 124 - 62 = 62 \\ 62 + 72 = 134 \end{array} \right.$
 _____ $\left\{ \begin{array}{l} 134 \div 2 = 52 \\ 124 - 72 = 72 \end{array} \right.$
 $124 - 72 = 52$

test cannot give so thorough a diagnosis as is yielded by the Compass series of tests.

Survey Tests in Arithmetic.—Survey tests have functions, just as do diagnostic tests. These include (1) indicating the total attainment of pupils; (2) yielding a basis for grouping pupils, if such grouping is desired; (3) indicating whether or not specific diagnostic tests should be administered, since survey tests reveal weaknesses or strengths in total skill; and (4) suggesting when and how much preventive teaching is needed, in the first teaching of a unit of instruction to avoid later failures. The following discussion will deal with survey tests in three groups: (a) reasoning tests, (b) tests of fundamentals, and (c) tests combining both reasoning exercises and problems in computation in one instrument.

Reasoning Tests in Arithmetic. *The Los Angeles Diagnostic Tests: Reasoning in Arithmetic.*¹—This test is diagnostic in name only. In the directions for using this instrument, it is specifically stated that the test was prepared for two purposes: (1) providing a standardized survey test in reasoning and (2) providing a method of determining the points of error of individual pupils in order that appropriate remedial and constructive activities may be instituted. An error-analysis chart is provided, together with statements as to the nature of each problem, which is a valuable aid in diagnosing group weaknesses. The instrument does not adequately diagnose the difficulties of individual pupils.

The test includes 54 reasoning problems. The type of exercises in the instrument is indicated by the following excerpt from the test.

22. The girls wanted some fruit to take on a picnic. They bought 2 dozen oranges at 50¢ a dozen and a basket of peaches for 40¢. How much did they spend for fruit?

Work Here

Answer

23. I bought 5 dozen eggs at 18¢ a dozen and gave the store keeper a dollar bill. How much change did I receive?

Answer

¹ Devised by Carolyn Armstrong and published by the California Test Bureau, Los Angeles.

Age and grade placement norms are available for the test.

This test is representative of the numerous similar instruments available. Naturally, the directions of different tests vary somewhat, as do also the specific problems and methods of presenting the exercises, but the general similarity is such among the several instruments that it is not necessary to provide further illustrations of reasoning tests.

Tests in Arithmetic Computation. *The Los Angeles Diagnostic Tests: Fundamentals of Arithmetic.*¹—This test, like the one discussed immediately above, is a survey instrument. It is intended for use in grades 2 through 8, and it provides a measure of the four fundamentals. There are 39 problems in the addition section of the instrument, 42 problems in the subtraction section, 41 problems in the multiplication section, and 45 problems in the division section. The exercises deal with whole numbers, fractions, decimals, and mixed numbers. There are a few exercises dealing with United States money, linear measure, and the like. The content of the instrument is well illustrated by the following excerpt from the test.

$$\begin{array}{r} (17) \\ 345 \\ 687 \\ 768 \\ 534 \\ 258 \\ \hline 874 \end{array}$$

$$\begin{array}{r} (18) \\ 7384 \\ 5053 \\ 4765 \\ \hline 8679 \end{array}$$

$$\begin{array}{r} (19) \\ \$.95 \\ 21.10 \\ 5.25 \\ .50 \\ \hline 14.75 \end{array}$$

(20)

$$\$6.25 + \$3.03 + \$8 + \$7.75 + \$5.49 + \$0.04 =$$

(21)

$$\begin{array}{l} 8 \text{ yd., } 1 \text{ ft., } 3 \text{ in.} \\ 3 \text{ yd., } 2 \text{ ft., } 4 \text{ in.} \\ \hline 1 \text{ yd., } 1 \text{ ft., } 5 \text{ in.} \end{array}$$

(22)

$$3 \text{ yr., } 8 \text{ mo.} + 4 \text{ yr., } 6 \text{ mo.} + 2 \text{ yr., } 8 \text{ mo.} =$$

Age and grade norms are available, as are tentative standards of accomplishment for each of the four fundamental processes by grades only. The Diagnostic Record Sheet provides an excellent error-analysis chart as well as an indication of the meaning of errors for each problem.

This instrument typifies in general survey tests of its type.

¹ Devised by Carolyn Armstrong and W. W. Clark and published by the California Test Bureau, Los Angeles.

Combination Tests. *The Spencer Diagnostic Arithmetic Tests.*¹ This is a series of three instruments. Test 1 is intended for grades 3 and 4; Test 2, for grades 5 and 6; Test 3, for grades 7 and 8. Tests 1 and 2 deal with the four fundamental processes, subtraction, addition, multiplication, and division. Test 1 includes problems dealing with whole numbers, United States money, weights and measures, and time. Test 2 deals mainly with numerical quantities, with exercises involving whole numbers, mixed numbers, fractions, and decimals. The materials of these two tests are sufficient to provide a reasonably adequate diagnosis. Test 3 includes 28 problems of a concrete type. On six of these, the pupil is asked only to indicate the correct way to solve each problem; and on the second six, he is asked, "What are you asked to find?" and "Which of these [given] methods will give the correct answer?" in addition to solving the problems. The last 16 problems are similar to those in the Los Angeles Reasoning Test in Arithmetic. It is doubtful that the materials in Test 3 are sufficiently extensive to provide an adequate diagnosis, and this is the reason that this test series is included among the survey instruments.

The following excerpts from Test 3 indicate the nature of the exercises dealing with problem analysis.

Place a cross (x) on the line which indicates the correct way to solve each problem.

1. If you know how much food one man should have and the number of men in a certain party, you can tell how much food will be needed for the party by

- adding the amount needed by one man to the number of men.
 multiplying the amount needed for one man by the number of men.
 dividing the number of men by the amount needed by each man.

5. A man paid \$.32 per hundred dollars for insuring his house against fire for one year. The policy was for \$5,000. What premium should he pay the company?

- | | | |
|--|---|--|
| (a) What are you asked to find? | } | The premium paid.....
The face of the policy.....
The amount of insurance..... |
| (b) Which of these methods will give the correct answer? | } | Divide \$5,000 by 100 and multiply by \$.32.....
Multiply \$5,000 by \$.32.....
Divide \$5,000 by \$.32..... |
| | | (c) The correct answer is..... |

¹ Devised by P. L. Spencer and published by the C. A. Gregory Company, Cincinnati, Ohio.

The scoring materials emphasize the diagnosis of pupil errors. Norms are not available. Despite this deficiency, this is a valuable instrument, particularly Test 3, because problem analysis is one aspect of arithmetic instruction that has not been sufficiently emphasized.

*The Schorling-Clark-Potter Arithmetic Test.*¹—This test is concerned largely with computation. It not only measures skills with whole numbers, but it also involves the various steps by which skills are built up in the less frequently measured subjects of common fractions, decimals, percentage, and denominate numbers. Part I includes 10 addition problems; Part II, 10 subtraction problems; Part III, 17 multiplication problems; Part IV, 16 division problems; Part V, 37 problems involving fractions, decimals, and percentages; and Part VI, 10 verbal problems.

Grade norms are available, both on total scores and for scores on the six parts of the test for two semesters each in grades 5 through 8. The results yielded by this test indicate the general status of pupils in the processes of arithmetic computation. Data from two testings can be used in measuring progress in arithmetic achievement in general.

*The Sangren-Reidy Survey Tests in Arithmetic.*²—The purposes of these instruments are (1) to provide a measure of the general level of pupils' ability in the major processes of arithmetic and (2) to determine the phases of arithmetic and to locate the level of difficulty in which the pupil needs further diagnostic testing in order to discover his instructional needs. The tests are constructed in three divisions: I, for grades 2 and 3; II, for grades 4, 5, and 6; and III, for grades 7 and 8.

Division I includes 14 problems in addition, 15 in subtraction, 13 in multiplication, and 12 in division. The problems in addition and subtraction include whole numbers and fractions, and those in multiplication and division, whole numbers only. Division II includes 13 problems in addition, 13 in subtraction, 19 in multiplication, 15 in division, and 15 reasoning problems. The exercises dealing with fundamental processes include whole numbers, fractions, mixed numbers, decimals, percentages,

¹ Devised by R. Schorling, J. R. Clark, and Mary A. Potter and published by the World Book Company, Yonkers-on-Hudson, N. Y.

² Devised by P. V. Sangren and Ann Reidy and published by the Public School Publishing Company, Bloomington, Ill.

materials, rather than to examine critically any large amount of the excellent test material that is available in arithmetic. The student is urged to study materials listed for further reading and particularly to examine the test content and test manuals of a number of standardized tests in the subject.

Readings for Further Study

1. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. VII.
2. W. S. MONROE, J. C. DEVOSS, and F. J. KELLY, *Educational Tests and Measurements* (rev. ed.). Chap. II.
3. S. L. and L. C. PRESSEY, *Introduction to the Use of Standardized Tests* (rev. ed.). Chap. V.
4. H. B. REED, *Psychology of Elementary School Subjects*. Chaps. XV-XVIII, inclusive.
5. H. L. SMITH and W. W. WRIGHT, *Tests and Measurements*. Chap. V.
6. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. IX.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. IV.

Questions and Exercises

1. Choose one arithmetic test particularly useful in diagnostic measurement, and study it carefully. What characteristics make it particularly valuable for analytical measurement?

2. What standardized tests are available to aid a teacher in determining whether or not her pupils are deficient in arithmetic vocabulary?

3. How would a teacher proceed, using standardized tests, in determining whether too much or too little time was being given to arithmetic instruction?

4. Is a single-class median score that is conspicuously above the norm for end-of-semester testing in arithmetic an indication of superior teaching? Why?

5. What reading skills enter largely as factors in the success of pupils in arithmetic reasoning? Why?

6. What are the two major problems in measurement in arithmetic?

7. Which standardized arithmetic test or tests should a principal choose for making a general survey of the arithmetic achievement of the pupils in grades 4 to 6 inclusive? Why? Can the results of such a survey be used in analyzing class weaknesses? If so, how?

8. Select three standardized tests of arithmetic computation skills and three such tests of arithmetic reasoning skills. What are the outstanding excellences of each test? The significant contributions, if any, of each test to measurement techniques?

9. Describe techniques used in standardized tests for diagnosing pupil difficulties in solving problems.

10. What assistance can a teacher obtain from testing procedures in determining whether or not the methods she is using in her classroom are successful?

11. What is the purpose of speed or rapidity of work in arithmetic computation? Do the standardized tests measure this variable? Illustrate by reference to standardized tests.

12. Are the standardized tests of arithmetic reasoning skills contained in batteries of achievement tests entirely adequate? Why?

13. Is it easier or more difficult to standardize arithmetic tests than tests in other school subjects? Why?

14. Outline a procedure to be followed by a principal of an elementary school in testing arithmetic achievement for a period of one school year, dealing with all grades and assuming that pupils in grades 4 and 5 are weak in computation.

15. Compare the tests of arithmetic fundamentals contained in batteries of school achievement tests with survey computation tests. What similarities and what differences are found?

16. Develop two arithmetic reasoning exercises, such as might be included in a test for grade-8 pupils, covering (1) reading comprehension, (2) analysis and theory, (3) procedure in solving the problem, and (4) the computation and the answer.

CHAPTER IX

STANDARDIZED TESTS IN THE SOCIAL STUDIES

Social science has had a place in the curriculum of elementary schools for many decades, but it has been largely presented as a group of separate subjects, such as history, geography, and civics. During recent years, a tendency has developed to discontinue these separate and specialized fields of the social sciences as subjects of instruction in elementary schools and to substitute in their place a composite *social science* course, drawing its content from a number of phases of the social sciences.

Social science definitely has been made responsible for training pupils for citizenship. The emphasis in present-day instruction is upon the functional interpretation of subject matter rather than upon the mere mastery of facts. The facts in social science are numerous, and many of these, such as the facts of geography and history, the locations of cities, and the dates of historical events, are well established in social science textbooks designed for use in elementary schools. Pupils have been, but should not be, expected to learn all of these facts. No elementary-school pupils could learn all of the content of the social studies; the material of social science is life as it has been and now is lived by men, and it is constantly and rapidly changing.

The shift in emphasis has been from the memorization of the facts of geography, of history, or of civics to the effort to develop adequate habits of thinking about social problems, based on accurate information, and to guide pupils in acquiring proper social attitudes which will result in their later engaging in socially desirable behavior. The effort is to give pupils information and skills and attitudes that will be of value to them throughout their entire lives, in dealing with the incidents of social life as they are encountered.

As in other school subjects, good tests are necessary to good teaching in the social studies. The tests now available concentrate largely upon the measurement of specific facts; they yield

little data as to pupils' attainment of the dynamic values of instruction in the social studies. It is true that there is definite justification for the precise testing of exact information that pupils have been required to learn, whether this information includes specific places, persons, events, symbols, definitions, or what not, but such testing of precise information does not at all cover the scope of what should be tested in the teaching of social science.

Among the skills essential in the social sciences that are not measured by standardized tests in social science, devised for use in elementary schools, are note taking, requiring the comprehension and analysis of material read; the preparation of maps, which is a fundamental skill in both geography and history; the preparation and reading of graphical materials; the understanding of ideas, which is a more important result of social science instruction than is mastery of any body of specific information; and the ability to analyze critically or to appraise information. The available standardized tests do not measure interests and attitudes adequately. The conclusion is that they are inadequate for the complete and accurate measurement of achievement in the social studies, even though they may measure factual information quite well.

It seems that the improvement of standardized tests in the social sciences must wait for curricular developments in this subject field as well as improvements in methods of instruction. When these things have been accomplished, it will be possible to standardize adequate tests to measure attainment in the social studies among elementary-school pupils.

Despite their inadequacies, it is necessary for the classroom teacher to know the character and the content of existing tests that have been devised for use in the social sciences in elementary schools. The following discussion will deal with these instruments in this order: (1) standardized tests of geography; (2) standardized tests of history; and (3) standardized tests of civics.

Tests in Geography. *The Wiedfeld-Walther Geography Test.*¹—Geography tests ordinarily measure only information, perhaps because of the ease with which this aspect of geography can be measured. The Wiedfeld-Walther test attempts to accomplish

¹ Devised by M. Theresa Wiedfeld and E. C. Walther and published by the World Book Company, Yonkers-on-Hudson, N. Y.

the measurement of geographical information and also the measurement of skills in finding, evaluating, and using of facts in geography.

This instrument is designed to provide an inventory of pupils' achievements in the several phases of geography usually taught in grades 4 through 8. The test contains two parts: Part 1, which measures the abilities and skills needed in the study of geography; and Part 2, which measures knowledge of geographical facts. Part 1 tests the extent to which three necessary study abilities and skills have become part of a pupil's thinking and behavior in the effective study of geography. This portion of the instrument includes three tests: Test I measures the ability to read various types of geographical material and to comprehend their meaning for use in geographical interpretation. Test II measures the ability to recognize the worth of geographical material with relation to geographical problems. This test requires pupils to organize ideas with relation to a problem showing relationships of (1) geographical cause and effect, (2) general and specific statements of geographical fact, and (3) main and subordinate ideas. The problems are limited in scope, and they are simple enough for the relationships to be evident at the first reading. Test III measures the ability to do map and graph reading. Map reading is a complicated skill which means getting facts as to topography, land forms, climate, production of goods, and the like, that may be represented in a map. Graph reading is a skill essential to the attainment of geographical facts and relationships. This test measures these abilities to use these special tools of geography for specific purposes.

Part 2 measures the memory control that pupils have over three of the phases of geographical knowledge most frequently listed among goals of achievement in courses of study and most necessary to attainment in geography. This portion of the instrument also includes three tests. Test IV measures pupils' knowledge of word associations built up in discussions and readings. The words in this test represent the vocabulary of (1) physical geography, dealing with such items as land and water forms, characteristic vegetation, animal life, and climate; (2) social geography, dealing with such items as peoples, dress, transportation, homes, and religion; and (3) political geography, dealing with items such as government and currency. The words

are distributed as to regions. Test V measures pupils' knowledge of geographical factors and human response to them. It deals with economic, social, political, and physical geography representative of different regions and different industries. Test VI measures locational geography, involving pupils' knowledge of places and their locations necessary to an intelligent understanding of national and world happenings and conditions.

A portion of Test I is given below to indicate the type of content included in the test.

TEST I. READING

DIRECTIONS. Here are some paragraphs just like the paragraphs you read in your geography lessons. The sentences in each paragraph are numbered. Each paragraph is followed by a question. *One* sentence in the paragraph contains information that helps to answer the question better than any other sentence in the paragraph.

First read the paragraph. Then read the question. Then pick the one sentence that you think helps best to answer the question. Write the *number* of the sentence you choose in the parentheses after the question. The sample is done correctly for you.

SAMPLE.

¹ Because of the long, warm, moist summers of the South crops are raised which are not raised in the North. ² Fruits and vegetables are raised. ³ Thousands of bushels of many kinds of nuts are raised every year. ⁴ These products are sold in Northern markets.

Why is cotton raising an important industry of the Southern states? (1)

¹ People in a city eat enough eggs in one day to fill a small house. ² They buy eggs from stores. ³ The storekeepers buy them from farmers. ⁴ Most farmers raise large flocks of chickens.

1. Why do so many farmers raise chickens?.....(1)

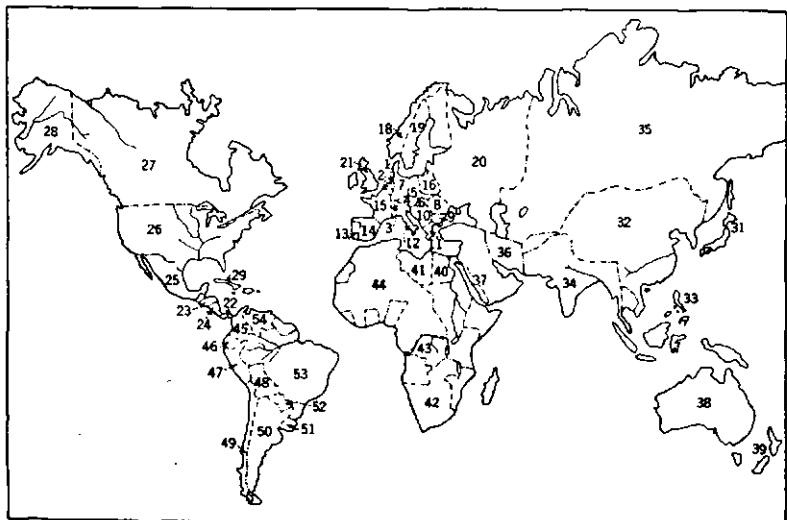
This somewhat unusual geography test is useful in determining the specific difficulties of pupils studying the subject and thus in providing the teacher with information leading to directed instruction, drill, and review aimed at the abilities and needs of the class and of individual pupils. The content of the test is perhaps somewhat difficult, since the instrument involves

thought and reasoning beyond the mere recall of factual information.

*The Gregory-Spencer Geography Tests.*¹—This instrument was devised for use in grades 6, 7, and 8. It is a companion test to the Gregory-Hagerty Geography Tests, which are intended for use in grades 4, 5, and 6.²

The Gregory-Spencer Geography Tests include three duplicate tests. Each includes eight parts. Part 1 deals with trade routes and their products; Part 2, with miscellaneous geography; Parts 3 and 4, with casual geography of the United States and of the world; Parts 5 and 6, with place and descriptive geography; and Parts 7 and 8, with political and place geography. Parts 7 and 8 are reproduced below, to provide an idea as to the content of the examination.

PARTS 7 AND 8—POLITICAL AND PLACE GEOGRAPHY



PART 7. In the left-hand column of the table [on page 230] is a list of 11 countries. Each of these countries is represented by a number on the map. Find the number on the map that represents each of the countries and write it below in the blank space to the left of the country. Example: Canada is the first country mentioned in the list [on p. 230]. Looking on the map we see that Canada is represented by the number 27, so we have written the number 27 in the blank space to the left of the word, Canada. Put the numbers that represent each of the other countries in the blank spaces in the same way. **DO THIS PART OF THE TEST BEFORE YOU READ THE DIRECTIONS FOR THE NEXT PART.**

¹ Devised by C. A. Gregory and P. L. Spencer and published by the C. A. Gregory Company, Cincinnati, Ohio.

² Devised by C. A. Gregory and Mary E. Hagerty and published by the C. A. Gregory Company, Cincinnati, Ohio.

PART 8. In the middle of the page below is a list of 11 descriptive phrases. Each of these phrases describes one of the 11 countries in the left-hand column below. Read the phrase and write the name of the country that it describes in the blank space to the right in the column marked "countries." Example: the sixth phrase from the top reads, "Largest British possession in America." This phrase describes Canada, so we have written the word, "Canada," in the blank space at the right of the phrase. Read each of the phrases in the order they are given, starting at the top and write the name of the country which each describes just as we have done for the phrase which describes Canada.

No.	Countries	DESCRIPTIVE PHRASES	Countries
27	Canada	Wealthiest nation in the world, second only to England in size of merchant marine.	
	Egypt	Most progressive country in the East; sometimes called the "England of the Orient."	
	Brazil	A fertile agricultural country due to the overflowing of the Nile river, and famous for its museums, sphinx and pyramids.	
	India	Largest country in Europe. Ranks next to the United States in the production of grain and petroleum.	
	United States	Largest country in South America. Important in the production of rubber and coffee.	
	Russia	Largest British possession in America.	Canada.
	Japan	The country which leads the world in the building of ships and the manufacture of fine cutlery, woolen and cotton goods.	
	British Isles	North American country crossed about mid-way by the Tropic of Cancer. The home of the once famous Aztec Indians. Now noted for its minerals and petroleum.	
	Norway	Asiatic country sometimes called "Great Britain's Granary." Important for production of tea, grain, cotton and other raw materials.	
	France	Scandinavian country which is so barren and poorly adapted to agriculture that the people are forced to the sea for most of their food.	
	Mexico	European country largely engaged in agriculture, but because of the artistic tastes of its people is also very important in the manufacture of the best grades of pottery, textiles, etc.	

Part 7.—Score = number right.....

Part 8.—Score = number right.....

The separate parts of this test are quite short, and it seems doubtful that the instrument will yield an adequate measure of total geographical knowledge over a three-grade range. The test is open to criticism as to its format. The length of line, the narrow leading of parts of the test, and the small, heavy-face type make it rather difficult to read. The norms are inadequate, with only medians for end-of-year achievement given. In spite of these defects, this is one of the better known and more often used tests of geography achievement.

History Tests. *The Iowa General Information Test in American History.*¹—This test, devised for use in grades 7 and 8, includes 100 incomplete statements of facts about American history. The

¹ Devised by M. H. DeGraff, G. M. Ruch, and H. A. Greene and published by the Bureau of Educational Research and Service, State University of Iowa, Iowa City.

type of content is illustrated by the following excerpt from the test:

- 46 To establish a home for the poor and indebted classes of England, a great philanthropist established in America the colony of..... (46) _____
- 47 The two greatest rival Italian cities carrying on extensive trade with the East prior to 1450 were Genoa and (47) _____
- 48 Columbus' ships were equipped by money obtained from the sale of jewels of..... (48) _____
- 49 "I propose to fight it out on this line if it takes all summer," said..... (49) _____
- 50 The constitutional amendment which prohibits slavery or involuntary servitude except as punishment for crime is amendment number..... (50) _____

This test lacks the scope of content possessed by certain other tests, but it covers typical materials, presented in objective form.

*The Denny-Nelson American History Test.*¹—This test was devised for use in grades 7 and 8. The test includes two parts: Part I, covering the period of American history of discovery, exploration, and colonization, ending with 1789; and Part II, covering the National Period, from 1789 to the present. Both parts include a few items concerning the Constitution. Part I

F. DIRECTIONS. The name of each person below is given a number.

Write the number of the name in the parentheses after the statement that suggests the person's name.

- | | |
|--------------------------|---|
| Cyrus H. McCormick (1) | 76. First Chief Justice..... () |
| Brigham Young (2) | 77. Negro education..... () |
| Alexander Hamilton (3) | 78. Acquitted when tried for treason..... () |
| John Jay (4) | 79. Louisiana Purchase..... () |
| Henry Clay (5) | 80. The Mormons..... () |
| James Monroe (6) | 81. Invention of the cotton gin..... () |
| George Dewey (7) | 82. Raid on Harper's Ferry..... () |
| Booker T. Washington (8) | 83. Invention of the reaper..... () |
| Samuel F. B. Morse (9) | 84. First Secretary of the Treasury..... () |
| Thomas Jefferson (10) | 85. Compromise of 1850..... () |
| Eli Whitney (11) | 86. Editor of the "Liberator"..... () |
| John Brown (12) | 87. Invention of the telephone..... () |
| A. G. Bell (13) | 88. Invention of the telegraph..... () |
| Aaron Burr (14) | 89. Era of Good Feeling..... () |
| William L. Garrison (15) | 90. Battle of Manila Bay..... () |

¹ Devised by E. C. Denny and M. J. Nelson and published by the World Book Company, Yonkers-on-Hudson, N. Y.

contains five different types of questions, with 80 items in all. Part II contains seven different types of questions, with 100 items in all. The test items are varied in character. The content on page 231, the sixth section of Part II, illustrates the type of materials included in the test.

This is one of the better American history tests available for classroom use.

Civics Tests. *The Burton Civics Test.*¹—This instrument was devised to function as an inventory test of pupils' general knowledge of civics—the political, economic, and social facts of life. It is useful in grades 5 through 8 in the elementary school. The instrument is quite brief, including only 60 items. Content is illustrated by the following three items excerpted from it.

14. What is impeachment?.....
1. Accusing a public officer of improper conduct
 2. Calling a man to serve on the jury
 3. Spending more money than one should
15. What is a consul?.....
1. A lawyer who gives counsel or advice
 2. An official representing his government in foreign lands
 3. A doctor who is called into consultation
16. What is a city ordinance?.....
1. A city officer
 2. A city law
 3. A city map

The weakness of this test is its brevity and the fact that it measures information only. No attempt has been made in it to measure appreciations, attitudes, or judgments; this, however, is a weakness common to tests in the social sciences.

*The Brown-Woody Civics Test.*²—This instrument was devised primarily for use in high-school grades, but it has norms also for grades 7 and 8 in the elementary school. The content requires an understanding of the vocabulary of civics, a mastery of the facts of civics, and the application of the principles of civics to practical situations.

Part I of this test measures the pupils' mastery of the vocabulary of civics; Part II, the extent of their information in civics; and Part III, the extent of their ability to do civic thinking.

¹ Devised by W. H. Burton and Virginia N. Burton and published by the World Book Company, Yonkers-on-Hudson, N. Y.

² Devised by A. W. Brown and C. Woody and published by the World Book Company, Yonkers-on-Hudson, N. Y.

This portion of the test also permits the examiner to determine whether or not the pupils possess proper civic attitudes, ideals, and appreciations. The example from the content of this test quoted below is chosen from Part III of the instrument.

DIRECTIONS. Imagine that you are a voter. There are two candidates seeking the office of mayor of your city. The qualities and characteristics of each are listed below. We shall call one Mr. A and the other Mr. B. You are to select the candidate who is, in your opinion, best fitted to perform the duties of that office. Read carefully the characteristics and qualities of each before you make your decision. Place an X in the square following the name of the candidate for whom you would vote.

Begin here.

Mr. A <input type="checkbox"/>	Mr. B <input type="checkbox"/>
Legally qualified for office	Legally qualified for office
Relative of Governor of the State	Experienced as a lawyer
Ex-senator	Previously an alderman of his ward
University graduate	Junior College graduate
Member of Board of Commerce	Member of Board of Commerce
President of an industrial corporation	Successful business man in the community
Member of several fraternal orders	Interested in city's welfare
"Friend of the laboring man"	Far-sighted
Wealthy	Conscientious
Critical of city management	Honest in dealing
Well known in social circles	Prominent in business circles
Church member	Sterling character
Fluent speaker	"Man of few words"
Conservative	Progressive
Friendly	Intellectual
Efficient politician	Organizer

After you have selected one of the above candidates, place a check mark (✓) before the qualities or characteristics that determined your choice.

This is the most valuable of the existing tests for the measurement of achievement in civics.

Summary and Conclusions.—In the opening paragraphs of this chapter, it was noted that history, geography, and civics have been taught as separate subjects in the past and that, during recent decades, there has come a change in content and in methods of teaching in the social studies. The social studies have been made responsible for training pupils for citizenship, and there has been a trend away from the mere memorization of facts to the

effort to develop adequate habits of thinking about social problems, based on accurate information, and proper attitudes and ideals of citizenship.

It should be noted in the test materials presented and described in this chapter that these instruments measure primarily factual information. The ability to evaluate critically—to reason—is measured to a small amount by certain of the standardized tests. They do not measure interests and attitudes adequately.

Further research is needed in the measurement of the social sciences, and new and better standardized tests must be developed. These tests should be delayed until curricular practice in the social studies has been somewhat stabilized, and teaching practices improved. Whatever their character, these new tests should provide survey instruments that will measure all phases of the achievement of pupils in the social studies and will indicate whether or not the social sciences actually are accomplishing their objectives. Until this research is accomplished, and these new instruments are available, teachers should continue (1) to use the available standardized tests, taking advantage of the values that they have; and (2) to develop their own testing techniques and testing procedures to a high level to do as precise and thorough a job of measurement as may be done.

Readings for Further Study

1. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chap. XII.
2. T. L. KELLEY and A. C. KREY, *Tests and Measurements in the Social Sciences*. Part IV: Report of the Commission on the Social Studies, American Historical Association. New York: Charles Scribner's Sons, 1934.
3. E. M. PAULU, *Diagnostic Testing and Remedial Teaching*. Chaps. XI, XII.
4. G. M. RUCH and G. D. STODDARD, *Tests and Measurements in High School Instruction*. Chap. IX.
5. H. L. SMITH and W. W. WRIGHT, *Tests and Measurements*. Chaps. X, XI.
6. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chaps. XIV, XV.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chaps. XII, XIII.

Questions and Exercises

1. Compare and contrast the content of social science tests in the Public School Achievement Tests, the Unit Scales of Attainment, the Metropolitan

Achievement Tests, and the New Stanford Achievement Tests with that of two standardized tests in each of the following subject fields: civics, geography, and history.

2. Discuss the advantages of a measuring instrument such as the Gregory-Spencer Geography Tests over an instrument such as the Hahn-Lackey Geography Scale. In what teaching situations might the scale be of greater value than the test?

3. Why are standardized tests in geography less important than are those in arithmetic and in reading?

4. Differentiate between a thought question and a memory question in geography. Illustrate each type.

5. What criticisms may be directed against the maps used in standardized geography tests? Why?

6. What effects will the unification of the social sciences for instructional purposes have upon testing procedures in geography and history?

7. Is it likely that standardized tests in geography and in history will ultimately be available as adequately standardized as the best of the current tests in reading and in arithmetic? Why?

8. Why is it still relatively difficult to standardize a history test? What must be done before the difficulties can be avoided?

9. Should civics materials be included in the content of a standardized American history test? Why?

10. The measurement of attitudes, *e.g.*, civic attitudes, is quite difficult. Why is this?

11. Name several outstanding American history tests. Evaluate them, listing their excellences and their defects.

12. Is it possible to make an inventory test in history? Why?

13. Examine several tests of American history. Do they measure adequately recent American history? Justify your answer.

14. What practical objections may be raised against the use of the Ely-King History Tests (or any similar battery)? Would unit tests be preferable to standardized subject-matter end-of-semester tests in all instances?

15. What influences have recent developments in teaching methods had upon increasing or decreasing the emphasis upon fact knowledge as an outcome of instruction in the social studies? How will this affect testing procedures hereafter?

16. Assume that you are meeting for the first time a class in geography. Indicate what testing procedures you would follow.

CHAPTER X

STANDARDIZED TESTS IN OTHER ELEMENTARY-SCHOOL SUBJECTS

The discussion in the preceding three chapters has dealt with the standardized tests that are available for use in language arts, arithmetic, and the social studies in the elementary school. There are other subjects that are taught in elementary schools, agriculture, art, foreign languages, home economics, industrial arts, music, physical education, and science. Of these, only art, music, and science will be considered briefly in this chapter, since only these are commonly taught in schools throughout the country.

Art in the Elementary School.—This subject in the elementary-school curriculum is taught primarily because of (1) its social and leisure-time values and (2) its vocational value. In general, art in the elementary school is cultural, providing experience with aesthetic materials and aesthetic values for everyday use. That now taught is general art, and this functions vocationally only in providing an opportunity for exploration and a basis for further study. General art functions for the average individual through providing the ability to observe the art qualities in materials and to appreciate these materials for purposes of enjoyment.

In general, art tests may be divided into three classes: (1) free-hand drawing scales; (2) art ability tests; and (3) art appreciation tests. One example of each type will be discussed hereafter.

*The Kline-Carey Measuring Scale for Freehand Drawing.*¹—This is actually two groups of scales, each with four parts or separate scales. Part I is intended for use in all elementary grades. Each separate scale includes 14 specimens. The four scales of Part I deal with the drawing of a house, a rabbit, a human figure in action, and a tree. The drawing of the tree represents brushwork. The different specimens are assigned percentage values

¹ Devised by L. W. Kline and G. L. Carey and published by the Johns Hopkins Press, Baltimore, Md.

from 0 to 95, and each is accompanied by an analysis of the good and bad features of the drawing, together with suggestions for practice that should lead to improvement. These materials provide both the teacher and the pupil with a guide for some diagnosis of pupil production in freehand drawing as well as a guide to improvement. Naturally, better results will be obtained if the pupils' drawings are of one of the subjects treated in Part I. This scale has been criticized because of its form and because of certain statistical matters concerning the standardization of the instrument. Nevertheless, it is probably the most useful instrument of its type now available.

Part II includes four parts or separate scales, each including 10 specimen drawings. The themes are illustrations, posters, borders, and structural design. This also is intended for use in all elementary-school grades. Factors of design and composition provide the criteria utilized in scoring, although the aesthetic principles of balance, emphasis, rhythm, opposition, and unity are also involved. Despite the shortcomings of scales of this type, they do serve to provide a means of improving the technique of art education in elementary schools.

*The Lewerenz Tests in Fundamental Abilities of Visual Art.*¹—These tests are three in number, devised for use in measuring seven phases of general art ability in nine subtests. Part I includes tests of (1) recognition of proportion and (2) originality of line drawing; Part II, tests of (3) observation of light and shade, (4) knowledge of subject-matter vocabulary, and (5) visual memory of proportion; and Part III, tests of (6) analysis of problems in cylindrical perspective, (7) analysis of problems in parallel perspective, (8) analysis of problems in angular perspective, and (9) recognition of color. It should be noted that Tests 6, 7, and 8 involve one phase of art production—analysis of problems in perspective.

This series of tests attempts to measure abilities rather than the products of abilities; *i.e.*, it attempts to ascertain the abilities possessed by pupils that will lead to success in art, and in this sense, the instrument is predictive of achievement rather than being a measure of achievement itself. It has adequate norms, and a profile chart is available.

¹ Devised by A. S. Lewerenz and published by the California Test Bureau, Los Angeles.

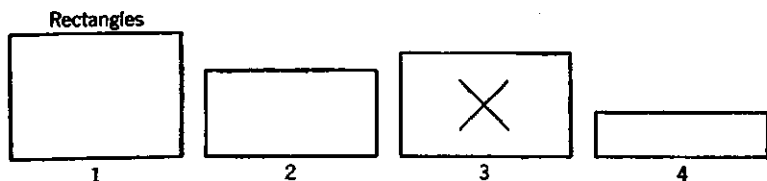
The type of content included in this instrument is indicated by the materials given below, chosen from Test 1.

DIRECTIONS: (*To be read aloud by examiner and silently by pupils*).

This is a test to show how well you can judge designs and shapes. On the two pages are fifteen sets of pictures. Each set includes figures from bad to good in shape. You are to pick out the one you like best in each set and mark it with an (X). The first set is a sample. Rectangle Number 3 is the best one so it has been marked with an (X). Now do the other fourteen sets in the same way, beginning with the bowls.

(*Time Limit: 10 minutes*)

(Sample Set)



*The Meier-Seashore Art Judgment Test.*¹—The two tests previously discussed in this chapter dealt with the production of art materials, but this instrument is concerned with the measurement of certain important aspects of aesthetic judgment or insight. This is only one of the many qualities underlying artistic talent, but it is of significance for eventual success in the field of art. This instrument is primarily of importance to the student interested vocationally in art, since it indicates with dependable certainty the important attitude for self-criticism by means of which the individual attains a constantly improved technique and quality of production. Since some artistic ability must be developed before this test is of great value, the instrument probably should not be used generally below the seventh grade in the elementary school.

The test is practically self-administering. Taking it makes necessary the use of a booklet of pictures, arranged in pairs. The 125 pairs of pictures differ in each pair in only one respect, and the pupil taking the test is told what differs in each pair. The following sample portion illustrates the nature of the test content (excepting the picture materials).

¹ Devised by N. C. Meier and C. E. Seashore and published by the Bureau of Educational Research and Service, University of Iowa, Iowa City.

L	R	44	Distribution of clearness
L	R	45	Inclusion or omission of picture on wall
L	R	46	Arrangement of inner parts of design
L	R	47	Position in the picture of the mounted riders
L	R	48	Presence or absence of tall reeds
L	R	49	Character and arrangement of waves

The pupil compares the two pictures in each pair, noting the unlike portion, and then decides which one is more artistic or pleasing. The test is untimed. Norms are available. This is an excellent measure of aesthetic sensitivity, which is a significant index to talent and to probable success in art.

Before leaving a discussion of tests in art, certain comments should be made. Compared with the many tests in reading and in arithmetic, the list of art tests is limited. Artistic skill is composed of a hierarchy of abilities, with those of less importance superimposed upon more fundamental traits. These broader and more fundamental traits are not measured by present standardized tests in art. For example, art appreciation tests now available do not adequately evaluate the emotional elements involved in the appreciation of art materials. Research is needed before abilities in art can be completely and adequately measured.

Music in the Elementary School.—Like art, music is taught in the elementary-school grades because of (1) its social and leisure-time values and (2) its vocational value. The purpose of music instruction in schools is to develop a conception of music as a lovely essential in normal living rather than to develop musical proficiency. It is true that public-school music instruction does tend to discover musical talent early in life and thus functions vocationally. This early discovery is essential because of the long training necessary to perfect music techniques and to achieve mastery of musical literature. For the great majority of pupils, however, public-school music functions in developing some degree of music ability plus some capacity for music appreciation.

Music tests now available may be divided roughly into three groups: (1) tests of musical talent, (2) tests of musical accomplishment, and (3) tests of music appreciation. Prognostic tests are essential if pupils are to avoid waste of time and energy in training for a musical career for which they have little or no innate talent, and they are equally essential for the encouragement of pupils with music aptitude who find results coming slowly and are

becoming discouraged because of a lack of progress. Two examples of this type will be discussed hereafter. Tests of musical achievement are more numerous than other types of music tests, perhaps because they are relatively easy to construct and because the mechanical aspects of music are more easily measured than are either musical capacity or musical appreciation. Two tests of this type also will be discussed later on. Musical appreciation is subtle in its measurement, and few tests of this ability are available. One such instrument will be discussed subsequently.

*Tests of Musical Talent: The Seashore Tests of Musical Talent.*¹ These tests have been constructed to measure the intangible quality known as musical talent; they aim to discover innate capacity to profit by training in music. They consist of six phonograph records, each involved in measuring a single ability. The six things measured are (1) the sense of pitch, (2) the sense of intensity, (3) the sense of time, (4) the sense of consonance, (5) tonal memory, and (6) the sense of rhythm. The tasks may be illustrated by quoting the directions for testing the sense of pitch: "You will hear two tones which differ in pitch. You are to judge whether the second is higher or lower than the first. If the second is higher, record H; if lower, record L." The scoring is accomplished by means of a key, and the scores, in terms of the number of correct responses, are translated in percentiles for purposes of comparison with norms. The norms are percentiles, and they are available for grades 5 and 8.

A considerable amount of research concerning these tests has been attained. Apparently, they accomplish their purpose fairly well. It has not yet been finally established that individuals who make high scores will become proficient musicians, and yet those who earn low scores have little opportunity for achieving a career in music. The value of the prognosis is negative, then, rather than positive. Low scores, however, should not bar any pupil from the usual public-school music training, and certainly they do not imply that an individual cannot appreciate music.

¹ Devised by C. E. Seashore. The *Manual of Instructions* and the six phonograph records required in the test are furnished by the Columbia Gramophone Company, New York.

*The Drake Musical Memory Test.*¹—This is a single test which purports to measure musical capacity. It does not measure acquired information or progress in musical accomplishment; rather, it attempts to measure one of the fundamental abilities without which musical talent cannot function except in a mediocre manner—musical memory. The attempt is to measure musical talent in a broad sense, in the way that talent is recognized when it is observed in action in any form.

The test contains 12 exercises, of which the following is one:



The symbols S, K, T, and N mean: S, exactly the same melody; K, a change of key; T, a change in time; and N, a change of one or more notes. After two practice exercises, the examiner plays each melody on the piano twice, and the pupil being examined notes whether there are any changes in the key, the time, or any of the notes. The examiner calls each exercise by number.

Norms are available for pupils above seven years of age.

Like other prognostic tests of musical talent, this instrument appears to be of more value for predicting negatively rather than positively. Also, no proof is yet available that musical memory is the only fundamental ability necessary for successful musical performance. The test is a welcome addition to the list of available music tests; its value would be greatly enhanced if it could be shown that it actually does predict musical proficiency as well as or better than other existing tests devised for the same purpose that are more complex and more difficult in administration.


*Tests of Musical Achievement: The Kwalwasser-Ruch Test of Musical Accomplishment.*²—This is a group test, devised for use in grades 4 through 8 in the elementary school. The instrument

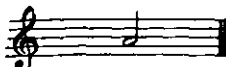
¹ Devised by R. M. Drake and published by the Public School Publishing Company, Bloomington, Ill.

² Devised by J. Kwalwasser and G. M. Ruch and published by the Bureau of Educational Research and Service, University of Iowa, Iowa City.

includes 10 parts measuring (1) knowledge of musical symbols and terms, (2) recognition of syllable names, (3) detection of pitch errors in a familiar melody, (4) detection of time errors in a familiar melody, (5) recognition of pitch names, (6) knowledge of time signatures, (7) knowledge of key signatures, (8) knowledge of note values, (9) knowledge of rest values, and (10) recognition of familiar melodies from notation.

The content of this instrument may be illustrated by a portion of Test 6, measuring knowledge of time signatures.

1  The time signature is $\frac{2}{4}$ $\frac{3}{4}$ $\frac{4}{4}$ $\frac{3}{8}$ $\frac{6}{8}$ 1

2  The time signature is $\frac{2}{4}$ $\frac{3}{4}$ $\frac{4}{4}$ $\frac{3}{8}$ $\frac{6}{8}$ 2

This is one of the most comprehensive standardized tests of musical achievement. It includes a wide range of subject matter, yet pupils can be expected to master the material in public-school music courses. The instrument provides an inventory of the knowledge and musical skill that a pupil possesses. Perhaps because the test is intended to cover high-school as well as elementary-school grades, the materials tend to be difficult for grade-4 pupils. Norms are available.

*The Knuth Achievement Tests in Music.*¹—This instrument attempts to measure pupils' ability to recognize and comprehend music from its notation. This is accomplished through measuring complete musical phrases rather than the dissected elements of music. No attempt is made to test isolated facts about music; music is considered as an integrated unit.

The test is composed of three divisions: (1) for grades 3 and 4, (2) for grades 5 and 6, and (3) for grades 7 and 8. Each division contains 40 items.

The presentation to the pupil is through the medium of the piano. A chord gives the signal for each item. Then a musical phrase of four measures is played. The pupil's test contains the first two measures as played, and his task is to select the last two from four printed musical illustrations. Testing requires the

¹ Devised by W. E. Knuth and published by the Educational Test Bureau, Inc., Minneapolis, Minn.

pianist to play the items at the exact metronomic tempo indicated and without error.

The test measures what it purports to measure. It is based upon recognition rather than performance, providing a knowledge of the pupil's recognition of the score apart from his ability in musical performance. It yields information for remedial teaching, although it is a little awkward to sort out pupils' errors.

*A Test of Musical Appreciation: The Kwalwasser Test of Music Information and Appreciation.*¹—This test was devised for use in high-school classes, and it is therefore too difficult for the great majority of upper grade elementary-school pupils. Only those with a considerable amount of training can respond successfully to this test. It is discussed here because it is the best available instrument of its type.

The test is presented in three divisions: history and biography, instrumentation, and musical form. The history and biography division includes five tests dealing with (1) classification of artists, (2) nationality of composers, (3) composers of famous compositions, (4) classification of composers by types of compositions, and (5) general knowledge of composers and compositions. The instrumentation section includes three tests which measure (1) production of tones on orchestral instruments, (2) classification of orchestral instruments, and (3) general knowledge of instrumentation. The musical-form division includes a single test which measures general knowledge of music structure and form. The illustrative material given below is chosen from this division of the instrument.

41. () The *sarabande* is a dance form.
42. () A *chanty* is a sailor song.
43. () *Recitatives* are found in operas and oratorios.
44. () The *concerto* is built on the rondo form.
45. () Beethoven substituted the *rondo* for the *scherzo* in the sonata.

It is obvious from the brief discussion that this test measures information more than it measures appreciation. It should not be forgotten that information is necessary for proper appreciation, even though it certainly is not all that is involved in appreciation.

If, as has been suggested, a sound philosophy of education emphasizes the intelligent consumption of music by the many,

¹ Devised by J. Kwalwasser and published by the Bureau of Educational Research and Service, University of Iowa, Iowa City.

and successful musical composition and performance at high levels of excellence by the relatively few, then it would seem that measurement in music has proceeded in an improper manner. More tests of musical information and appreciation should be available, and fewer of musical achievement. There is need for more research to accomplish the production of needed tests as well as to perfect the prognostic tests now available.

The following discussion will leave the measurement of fine arts, in favor of a presentation dealing with measurement in elementary science, another little-stressed subject in the elementary-school curriculum.

Science in the Elementary School.—Science has had only a small place in the instruction program in elementary schools, and there is a lack of uniformity with respect to science instruction among the schools in which it is taught. Usually, elementary science takes the form of health education or physiology in the intermediate or upper elementary grades. In all instances, it is a nonlaboratory course. These factors make the measurement of science in the elementary school a difficult problem, and at present only a few standardized tests are available in this field.

*The Dvorak-Van Wagenen Analytical Scales of Attainment in Elementary Science.*¹—This instrument includes three separate tests, of which two are intended for use in elementary grades, one in grades 5 and 6; and the other, in grades 7 and 8. Each of the tests includes three divisions, dealing with (1) physiology and hygiene, (2) biological sciences, and (3) physical sciences. Each of the three divisions contains 40 items in multiple-response form, of which the following excerpt from the biological science section of the test for grades 5 and 6 is an illustration.

- | | | | | | | | | | |
|---|-----------------|------------------|-------------------|-------------------|---|---|---|---|----------|
| 26. A pistil is a part of | 1. a bird | 2. a flower | 3. an insect | | | | | | |
| | 4. a seed | 5. a leaf | - | - | - | - | - | - | 26. |
| 27. Hepatica is a kind of | 1. cloud | 2. crystal | 3. water- | | | | | | |
| | fowl | 4. spring-flower | 5. grass | - | - | - | - | - | 27. |
| 28. A corolla is a part of | 1. a flower | 2. a bird | 3. an insect | | | | | | |
| | 4. a seed | 5. a leaf | - | - | - | - | - | - | 28. |
| 29. The carrying of pollen from the anther of one flower to the stigma of another is called | 1. reproduction | 2. fertili- | | | | | | | |
| | zation | 3. pollination | 4. pasteurization | 5. photosynthesis | | | | | 29. |
| 30. Bacteria are | 1. insects | 2. worms | 3. flies | 4. moths | | | | | |
| | 5. plants | - | - | - | - | - | - | - | 30. |

¹ Devised by August Dvorak and M. J. Van Wagenen and published by the Educational Test Bureau, Inc., Minneapolis, Minn.

These tests measure the ranges of information possessed by pupils in the branches of science included in the instruments. Obviously, the tests will be difficult for many elementary-school pupils, whose study of science has been limited.

*The Gates-Strang Health Knowledge Test.*¹—The Elementary Tests, a part of this instrument, are intended for use in grades 3 through 8. Each test includes 60 items, similar to the illustration quoted below from the test.

23. Going to the dentist twice a year

- | | | |
|--|-------|---|
| a. Can prevent all tooth decay..... | _____ | a |
| b. Can prevent small decayed spots in the teeth from becoming big..... | _____ | b |
| c. Keeps all bacteria out of your mouth..... | _____ | c |
| d. Makes all teeth white and straight..... | _____ | d |
| e. Saves you the trouble of brushing your teeth..... | _____ | e |

The Elementary Tests emphasize the less technical aspects of health knowledge, together with aspects of safety education. It should be obvious that so short a test cannot measure adequately over a six-grade range; yet this instrument is the best of its type now available.

The discussions in this chapter have dealt with the measurement of abilities in art and in music and with the measurement of science information. These subjects have had much less emphasis in the elementary-school curriculum than have the language arts, arithmetic, and the social sciences. Research is needed, both as to the curriculum in these subject fields and as to specific testing procedures and devices, before adequate measuring devices are available in the fine arts and in science.

A Note on a Procedure for Studying Standardized Tests.—This chapter closes the portion of this volume that is devoted to the presentation of standardized tests and scales. Throughout the discussion, an effort has been made to present the various types of measuring instruments that are available in the different subject fields. Obviously, it would be impossible to discuss extensively each separate test. Partly for this reason, and also partly because this volume is intended for beginners in educational measurement, no effort has been made to present extensive criticisms of individual instruments or specific data as to the

¹ Devised by A. I. Gates and Ruth Strang and published by the Bureau of Publications, Teachers College, Columbia University, New York.

validity and reliability of the different tests individually. The attempt has been to give a general insight into measurements in the different elementary-school subjects of instruction. Those students of measurements who wish detailed information on specific tests can secure it easily by obtaining and studying a sample copy of the instrument, together with the manual of instructions that accompanies it. In addition to this, the careful student will investigate reports in professional periodicals relating to the test or tests in which he is interested. Much valuable information is contained in periodical articles that has not yet found its way into textbooks.

Readings for Further Study

Art.

1. MADALINE KINTER, *The Measurement of Artistic Abilities*. New York: Psychological Corporation, 1933.
2. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. XVII.
3. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. VIII.

Industrial Education.

4. L. V. NEWKIRK and H. A. GREENE, *Tests and Measurements in Industrial Education*. New York: John Wiley & Sons, Inc., 1935.

Music.

5. J. KWALWASSER, *Tests and Measurements in Music*. Boston: C. C. Birchard and Company, 1927.
6. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chap. XVI.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. XI.

Physical Education.

8. J. F. BOVARD and F. W. COZENS, *Tests and Measurements in Physical Education* (2d ed., rev.). Philadelphia: W. B. Saunders Company, 1938.
9. RUTH B. GLASSOW and MARION R. BROER, *Measuring Achievement in Physical Education*. Philadelphia: W. B. Saunders Company, 1938.
10. C. H. McCLOY, *Appraising Physical Status: Methods and Norms, University of Iowa Studies, new ser., No. 2, Studies in Child Welfare, No. 15*. Iowa City: University of Iowa, 1938.
11. N. P. NIELSON and F. W. COZENS, *Achievement Scales in Physical Education Activities: for Boys and Girls in Elementary and Junior High Schools*. New York: A. S. Barnes & Company, 1934.

Questions and Exercises

1. What is the present status of measurement of art appreciation? What standardized tests are available for this purpose? What criticisms may be directed against them?

2. What are the fundamental abilities in visual art? How successfully has Lewerenz measured them in his *Tests in Fundamental Abilities in Visual Art*? Answer specifically.

3. Prepare a rough classification of the available standardized tests in art, under these headings: (1) freehand drawing tests; (2) art ability tests; and (3) art appreciation tests. Can you justifiably include the drawing scales in such a classification? Why?

4. Make a critical study of one of the drawing scales. What is the function of such an instrument?

5. Secure at least three samples of drawings by elementary-school pupils, and score them, using the Kline-Carey Measuring Scale for Free-hand Drawing, Part I. List the excellences and the defects of each drawing in terms of the information given on the scale, thus justifying your scoring of each drawing.

6. How may an art teacher use tests to make her teaching more efficient?

7. Indicate the research that is needed in the field of measurement in art.

8. What are the essential features of musical talent? What standardized tests are available for use in predicting success in musical training and performance?

9. What tests are available for measuring musical vocabulary? Sight singing?

10. List the difficulties that now prevent the obtaining of an objective measurement of musical appreciation.

11. Are the norms on most standardized music tests as likely to be of value for diagnosis as are the norms on a test of arithmetic computation? Why?

12. List the most important problems yet remaining to be solved in measuring adequately and objectively musical talent and musical accomplishment.

13. Why are there not now more standardized tests in the field of music?

14. What recent contributions have been made in music tests and measurements?

15. Why has not more attention been given to the development of standardized tests of science at the elementary-school level? What difficulties are encountered in standardizing such an instrument?

16. Check the several batteries of school achievement tests for the presence in each of a science test. Compare the content of such tests as you find in the batteries with that of single standardized tests in elementary science.

A Test on Standardized Subject-matter Tests

Directions: Each of the following statements is either true or false. If it is true, place a check on the line before the letter T; if false, a check on the

line before the letter F. If you do not know which is the correct answer, skip the item, and continue with the next one. DO NOT GUESS! Work as rapidly and as accurately as you can.

- | | |
|-----------------|---|
| _____ T _____ F | 1. The Gray Standardized Oral Reading Check Tests are arranged in sets to be used successively after periods of remedial work. |
| _____ T _____ F | 2. The Haggerty Reading Examination, 23, is used as an intelligence test. |
| _____ T _____ F | 3. The Sangren-Woody Reading Test is composed of different sections, each of which tests a different phase of silent reading comprehension. |
| _____ T _____ F | 4. The Sangren-Woody Test is an individual test. |
| _____ T _____ F | 5. The Gates Silent Reading Tests measure three types of reading comprehension. |
| _____ T _____ F | 6. The Buswell-John Diagnostic Test in Arithmetic is used to determine faulty habits in arithmetic computation. |
| _____ T _____ F | 7. The Buswell-John Test has a division for measuring the five fundamental processes in arithmetic: addition, subtraction, multiplication, division, and square root. |
| _____ T _____ F | 8. The Compass Diagnostic Test of Arithmetic is an individual test. |
| _____ T _____ F | 9. Pupils begin work on the Los Angeles Diagnostic Test of Arithmetic on the last rather than on the first page of the test booklet. |
| _____ T _____ F | 10. The words in the Buckingham Extension of the Ayres Spelling Scale are arranged in groups of equal difficulty. |
| _____ T _____ F | 11. The Iowa Spelling Scale provides norms for comparing pupils' achievement with an average. |
| _____ T _____ F | 12. The Ayres Gettysburg Handwriting Scale is constructed primarily to measure speed and quality of handwriting. |
| _____ T _____ F | 13. The Freeman Chart for Diagnosing Handwriting is contained in a booklet of one page for each of the characteristics of handwriting. |
| _____ T _____ F | 14. The Ayres Gettysburg Handwriting Scale is a handwriting scale rather than a test. |
| _____ T _____ F | 15. One of the characteristics of the Wilson Language Error Test is the "Pupil's Guide to Correctness in Written Work" which is used in connection with the test. |
| _____ T _____ F | 16. The Briggs English Form Test is a proofreading test. |
| _____ T _____ F | 17. One of the disadvantages of the Gregory-Spencer Geography Test is poor topographical composition. |

- _____ T _____ F 18. The Diagnostic Chart of the Buswell-John Test is for the pupil's use.
- _____ T _____ F 19. The Los Angeles Diagnostic Test in Language contains materials measuring achievement in punctuation, capitalization, and language usage.
- _____ T _____ F 20. The Clark Letter Writing Test measures pupils' ability to write business and social letters.
- _____ T _____ F 21. One of the outstanding characteristics of the Gregory-Spencer Geography Test is a set of large colored maps.
- _____ T _____ F 22. The Hahn-Lackey Geography Scale is a list of several hundred questions that can be scored objectively.
- _____ T _____ F 23. The Gregory-Hagerty Geography Test is an individual test.
- _____ T _____ F 24. The Denny-Nelson Test in American History is a list of subjective questions.
- _____ T _____ F 25. The Hill Civics Test is a series of three different measuring instruments for use in grades 6 to 8.
- _____ T _____ F 26. The Ely-King American History Test is divided into two sections, the questions in which cover different historical periods but are similar in complexity.
- _____ T _____ F 27. The Seashore Test of Musical Talent is an individual oral test.
- _____ T _____ F 28. The Kwalwasser-Dykema Music Tests are claimed to predict musical accomplishment.
- _____ T _____ F 29. A knowledge of musical symbols is essential to success on the Kwalwasser-Ruch Test of Musical Accomplishment.
- _____ T _____ F 30. Art vocabulary as well as elements of visual art is measured by the Lewerenz test.
- _____ T _____ F 31. The Burton Civics Test measures civic thinking and so is valuable in diagnosing civic problems in the classroom.
- _____ T _____ F 32. One of the characteristics of the Brown-Woody Civics Test is that separate norms are given for groups of varying nationality and economic status.
- _____ T _____ F 33. The Burton Civics Test gives problems on three separate phases of civics: economic, social, and political.
- _____ T _____ F 34. The Lee-Clark Reading Readiness Test is an individual test.
- _____ T _____ F 35. The Drake Musical Memory Test is given by the use of phonograph records.
- _____ T _____ F 36. The Kline-Carey Drawing Scale is an objective test of art ability.

- _____ T _____ F 37. A measure of reading readiness is included in the materials devised for use with the Betts-Keystone telebinocular.
- _____ T _____ F 38. The ophthalmograph is a motion-picture camera, useful in diagnosing faulty mechanics in silent reading.
- _____ T _____ F 39. The Willing Scale for Measuring Written Composition is completely objective in scoring.
- _____ T _____ F 40. The Nassau County Supplement to the Hillegas Scale is an instrument for grading English compositions.
- _____ T _____ F 41. The Compass Survey Tests in Arithmetic are useful for diagnosing pupil errors in the four fundamentals in arithmetic computation.
- _____ T _____ F 42. The Progressive Achievement Tests include tests in nine subject-matter fields.
- _____ T _____ F 43. One of the advantages in using a standardized test is the existence of norms that can be used in evaluating the general character of a pupil's achievement.
- _____ T _____ F 44. In the Franseen Diagnostic Tests in Language, the items are grouped according to the type of errors that they locate.
- _____ T _____ F 45. One of the difficulties involved in constructing tests of art and music appreciation is that appreciation involves the varying factor of emotion.
- _____ T _____ F 46. The Torgerson-Fahnestock Music Test is a sight-singing test.
- _____ T _____ F 47. There are very few standardized art tests now available.
- _____ T _____ F 48. The Public School Achievement Test includes a test of literary information.
- _____ T _____ F 49. The Unit Scales of Attainment include a test of elementary science.
- _____ T _____ F 50. The Pressey Chart for Diagnosis of Illegibilities in Handwriting is a scaled instrument.

PART II

Directions: Place a plus (+) on the line at the left of each subject field in which there are diagnostic tests; a minus (-), for those in which there are no diagnostic tests. If the validity or the reliability of the available tests is not adequate, encircle the plus or the minus. Consider only tests derived for use in elementary-school grades.

- | | |
|---------------------------------|-----------------------------|
| _____ 1. Silent reading | _____ 5. Physical education |
| _____ 2. Oral reading | _____ 6. Art (creative) |
| _____ 3. Arithmetic computation | _____ 7. Art (appreciation) |
| _____ 4. Arithmetic reasoning | _____ 8. Handwriting |

- | | |
|------------------------------|------------------------------------|
| _____ 9. Geography | _____ 13. Nature study and science |
| _____ 10. Civics | _____ 14. English (composition) |
| _____ 11. History | _____ 15. English (literature) |
| _____ 12. Health and hygiene | |

PART III

Directions: The following statements are to be completed with one or more words. Supply in the blanks the word or words that will best complete the statement. If you do not know the answer, skip the item, and continue with the next one. DO NOT GUESS! Work as rapidly and as accurately as you can.

- _____ types of tests are included in the Gates Silent Reading Tests for grades 1 and 2. Their names, which indicate the type of reading ability that they measure, are:
(A number)
- The Detroit Word Recognition Test is a _____ test. It contains _____ pictures and a like number of words and phrases, each naming or describing one of the pictures.
- The Sangren-Woody Reading Test for use in grades _____ to _____ consists of _____ parts, each purporting to measure a distinct type of reading ability. The tests are:

4. In the New Stanford Achievement Test there are _____
(a number)
arithmetic tests. They are:
- _____ should follow diagnostic testing.
- The Compass Diagnostic Tests in Arithmetic include _____
(a number)
tests measuring _____ different arithmetic skills. The tests
(a number)
are devised for use in grades _____ to _____.
- The purpose of the Stone Narrative Reading Tests is to measure three types of comprehension corresponding to three essential elements of a good narrative, which are:
a.
b.
c.
- Spelling tests in two achievement-test batteries are given as untimed dictated sentences. The batteries are:
a.
b.
- The Freeman Chart for Diagnosing Faults in Handwriting can be used to score _____ characteristics of handwriting. They are:
(a number)

10. The Pressey-Seaton Diagnostic Tests in English Composition, for use in grades _____ to _____, cover the minimum essentials in

- a.
- b.
- c.
- d.

11. The Charters Diagnostic Language Tests include _____
(a number)

separate tests. They are:

12. The Witham New Series Geography Tests include three tests, dealing with (1) _____, (2) _____, and (3) _____.

13. The Ely-King Battery of History Tests includes _____
(a number)

tests, which measure:

14. The Pressey-Richards Test for Understanding American History includes four subtests, measuring

- a.
- b.
- c.
- d.

15. The Seashore Tests of Musical Talent include six tests. These are:

16. The presentation to pupils of the Knuth Achievement Tests in Music requires the use of a _____.

CHAPTER XI

INTELLIGENCE AND ITS MEASUREMENT

The endeavor to devise adequate instruments for the measurement of human intelligence has occupied the time and the efforts of research workers in psychology and in education to a considerable degree during recent decades. The work has progressed; a vast amount of quantitative data have been accumulated; and a considerable literature has been developed relating to the nature and the measurement of human mental ability. Theoretical considerations relating to intelligence and its functioning are so interwoven and interrelated with practical attempts at the measurement of mental ability that neither can be discussed adequately without some consideration's being given to the other. It is the purpose of this chapter to discuss both the theoretical and the technical aspects of the subject. First, however, it seems desirable to present a short history of the movement, in order that the student may understand something concerning the origin and the development of present instruments used for and current procedures used in the measurement of mental ability.

History.—Man apparently has for many centuries recognized the existence of some internal force influencing his behavior in adjusting to his environment. He has also recognized the fact that this force or influence varied from one individual to another. As early as 25 centuries ago, Greek writers were developing scientific conceptions of intelligence and of mental operations.¹ Human interest in intelligence, now recognized as the force influencing human behavior, has been persistent and active through all these centuries.

Despite many years of thought and of research—for the present movement of measuring intelligence has a long history, involving the work of many careful scientists and extensive research

¹ J. PETERSON, *Early Conceptions and Tests of Intelligence*, pp. 11ff., Yonkers-on-Hudson, N. Y.: World Book Company, 1925.

projects—the beginnings of practical testing are relatively recent. The movement for the development and practical use of mental tests has a very short history in view of the extent to which such tests are now used.

In the early psychological laboratories, men attempted experimentation that had for its aim the discovery of general maxims concerning human behavior. In the process of these investigations, individual differences were discovered, and in turn these were studied. Their study led directly to the development of mental tests. For example, Galton's *Inquiries into Human Faculty* (1883) presented the general problem of the measurement of individual differences; and he soon thereafter established a laboratory in the South Kensington Museum in London where anyone could respond to a series of physical and mental tests.

Mental testing, as it is known today, is the product largely of the convergence of four streams of scientific development: German, French, English, and American. To a large extent, the movement was given impetus and was influenced by the work of Wundt's students in the first psychological laboratory, which he established in Germany about sixty years ago. Wundt himself had little sympathy with the problems of individual differences and with the measurement of intelligence.

Among Wundt's German students who played an important part in the development of mental tests were Kraepelin, Oehr, Meumann, Stern, and Ebbinghaus. Ebbinghaus experimented with memory (1885) and then developed a number of individual tests, publishing in 1897 a paper on a new method of measuring the mental ability of school children. This consisted of several individual tests of memory, computation, and the first form of his completion test.

Kraepelin, who became a physician, inaugurated a series of valuable studies on the differences between the minds of normal and insane individuals. In addition to this research, he developed a number of tests that have since become standardized and well known in the field of physical and mental measurement. Among these are endurance of grip, discrimination of brightness and of dual cutaneous impressions, computations, and rote memory. Oehr expanded the work of Kraepelin. Stern worked largely in the field of differential psychology. He was the first

man to suggest the use of a ratio between life age and mental age as a mental quotient.

French psychology has made its contribution largely through the work of Binet and his coworkers. Binet and Simon, however, are indebted to Blin (1902) and to his pupil Damaye (1903) whose work preceded theirs in 1904, 1905, 1908, and 1911. Blin's work was defective in that his tests were in part subjective, were not arranged in order of difficulty, and were subject to the effects of training. DeSanctis in Italy (1906) presented a scheme of tests that were the first scale after the preliminary work of Binet.

Binet was a member of a commission composed of medical men, educators, and scientists, who, at the instance of the French Minister of Public Instruction, studied the problem of special classes in the public schools during 1904. With Simon, Binet tried out his tests in schools, and from this preliminary work came his first rough scale in 1905. It consisted of a series of 30 tests, arranged in order of difficulty but not grouped according to age. Scoring was in terms of whole or half credit or none.

Binet reported little of importance during the following two years; but in 1908, he announced the first age scale. This included 59 tests, grouped from three to eight tests per year for ages three to thirteen. The method of giving several of the tests was changed from the 1905 scale, and many new tests were introduced, including several depending upon school knowledge. It was in connection with this scale that Binet advanced his method of calculating mental age for the first time. This scale was criticized by several research workers in Europe and in the United States on the ground that individual tests were not properly placed and that certain parts of the scale were too easy and others too difficult. Taking account of these studies and of the recommendations based on them, Binet revised his scale again, publishing the revised form in 1911. This dropped certain tests dependent on scholastic achievement, such as reading and writing; and certain other tests dependent upon knowledge of the environment, such as age and days of the week. In addition, the age placement of certain tests was changed, and certain new tests were introduced. The arrangement of the scale was improved, with five tests for each age level, except four tests at age four. This was Binet's final contribution to the mental-

testing movement. It is perhaps worth noting that he was motivated largely by an interest in the classification of the mentally defective. Throughout the entire development of mental tests in France, the influence of medical psychology is noticeable.

Until recent years, the contribution of England has been largely in the field of statistical treatment of test data yielded by normal-minded children and adults. Galton did secure anthropometric and psychometric measurements (1879-1904), but his great service was in developing and applying statistical methods of analysis. This work was further developed by his pupil Pearson and by Thomson, Burt, and Spearman.

American psychologists have been largely responsible for bringing the movement to its present status. Cattell, who was a student of Wundt, seems to have been the first to use the term "mental tests." As early as 1890, he published material describing tests in use in his psychological laboratory at the University of Pennsylvania. His tests included sensory discrimination, rapidity of reaction, and memory span. Later, with Farrand (1896) at Columbia University, tests were devised and given that involved handwriting, visual acuity and color vision, auditory acuity and pitch discrimination, sensitivity of skin, perception of weight, sensitivity to pain, accuracy and steadiness of movement, reaction time, cancellation of A's, memory span, perception of time and space, length of a line previously drawn, after images, and mental imagery. These tests were all given to individuals of normal intelligence.

Among the early pioneers in the United States was Münsterberg (1891), who gave various tests to school children but reported no results. Jastrow (1890-1893) at the University of Wisconsin gave tests and established norms that were later used in tests conducted by him at the Columbian Exposition, Chicago, in 1893. Boas and Bolton (1891) of Clark University made certain anthropological measurements of 1,500 school children, publishing their results. Gilbert at Yale University made studies of mental tests (1893-1896) that he applied to several hundred school children. This early impetus, however, was allowed to lag, and it was not until almost two decades later that interest was aroused again, after the work of Binet.

Numerous revisions of Binet's scales have been prepared in the United States. Goddard (1911) was the first to intro-

duce Binet's tests into America. He translated the scale into English and made certain changes in the tests. Several later revisions of the Binet Scale have appeared, among them those by Huey (1912); Wallin (1912); Kuhlmann (1912 and 1922); Terman and Childs (1912), Terman (1916) and Terman and Merrill (1937); Winch (1913); Strong (1913); Yerkes, Bridges, and Hardwick (1915) and Yerkes and Foster (1923); Porteus and Hill (1920); and Herring (1922). In the scales devised by Yerkes and his collaborators and by Herring, the classification of the tests into age groups has been abandoned in favor of a "point scale." The tests are arranged in order of difficulty, and points of credit are allowed for each test successfully passed by the subject.

The development of individual tests was not the only contribution of psychologists in the United States. Performance tests, such as the Pintner-Patterson Performance Test (1917), a group of 15 tests chiefly of the form-board type, and the Kohs Block Design Test (1921) represent contributions to mental measurement. The unique contribution of American psychology, however, is the group mental test, a pencil and paper test permitting the testing of large groups of individuals simultaneously. One of the earliest and certainly the most influential of such tests was the Army Alpha Test, prepared in 1917 for measuring recruits of the United States Army. This was taken by about 2,000,000 men. As a result of this work as well as of previous research, many such tests have been developed and used, principally in schools. Whereas during the early experimentation with mental tests in the United States interest was largely on research with the feeble-minded, today that interest has been largely transferred to the normal-minded.

It is difficult to give any clear conception of the significance of much of the research that has been done in so short a compass, but enough has been said to show clearly that present mental tests have developed from three sources: from school examinations; from the interview, such as physician with patient or employer with applicant for employment; and from early tests devised during the early history of psychology, *e.g.*, tests of memory, attention, and sensory acuity. Modern tests are similar to early tests in many respects, but they have been greatly improved.

With this brief discussion of the history of mental testing completed, it seems desirable now to turn to a consideration of the nature of intelligence.

The Definition of Intelligence.—The term intelligence has been used extensively in describing the relative performance level of human beings and of animals. Despite this fact, there is no single definition with which all educationists and psychologists agree; a wide diversity exists in the concepts held. Many varied definitions have been proposed, but there is still no one universally accepted view. It is necessary, therefore, to present not one statement of the meaning of intelligence but rather several different concepts, together with critical comments concerning each one.

Probably the most favored definitions may be classified as biological in type. Such definitions emphasize the adjustment of the individual to his environment or to certain aspects of his environment. Stern has given one statement: "Intelligence is a general capacity of an individual consciously to adjust his thinking to new requirements: it is general mental adaptability to new problems and conditions of life."¹ This definition restricts intelligence to mental adaptations; it emphasizes the adjustments to "new" elements in the environment; and it considers intelligence as a general capacity rather than as a complex of specific capacities.

Porteus has objected to certain implications in Stern's statement of his definition, as well as to similar statements by other psychologists. He writes:

In purely psychological terms the highest function of the cortex is to enable the animal to make new adaptations to situations. . . . In thus defining the highest cortical function as capacity for new adaptations, we part company with the majority of psychologists whose accepted definition of intelligence is "general mental adaptability to a new situation." It is unwise to lay so much emphasis on the newness of the situation instead of on the adaptive response.²

For his definition, Porteus has condensed the capacity for adaptation into "progressive adaptability." The emphasis in

¹ W. STERN, *The Psychological Methods of Testing Intelligence* (translation by G. M. Whipple), p. 3, Baltimore: Warwick and York, Inc., 1914.

² F. WOOD JONES and S. D. PORTEUS, *The Matrix of the Mind*, p. 436, Honolulu, T. H.: The University of Hawaii Press Association, 1928.

this statement is on the adjustment of the individual to his environment, and there is provision for the mental control of such adaptations as well as an explanation for both racial and individual maturation.

It is perhaps worth while at this point to give a criticism of this type of definition. McFadden has written:

The various definitions of intelligence seem to boil down to the general meaning: the ability to (or the capacity for) adapting adequately to the environment, an adequate and efficient adaptation being either concrete or abstract in nature. Now this sort of definition really means nothing at all, for we have no criterion for adequacy of adaptation or, perhaps, we have too many different criteria, which amounts to the same thing.¹

Spearman comments concerning the biological definitions of intelligence . . . "psychologists have 'found a formula' with which everyone can agree—providing that each interprets it differently."²

There has been another much favored attempt at defining intelligence. This kind of definition, advocated by certain educationists, places emphasis on learning ability. Thus, Buckingham says that intelligence is the ability to learn.³ If we are to understand from this definition what intelligence is, we must first define learning, and we must then determine what mental processes function to produce it. To call intelligence the ability to learn may supply a valuable statement about intelligence or possibly an aspect or phase of intelligence, but it does not define specifically the nature of intelligence.

There are other modern concepts of intelligence that are neither biological nor educational in their implications. Thus, Terman writes that "an individual is intelligent in proportion as he is able to carry on abstract thinking."⁴ This definition restricts intelligence definitely to one type of capacity. Hull, on the contrary, defines intelligence as "a kind of average of all possible

¹J. H. McFADDEN, The "Will-o'-the-wisp" Intelligence, *Psychological Review*, 39: 225 (1932).

²C. SPEARMAN, *The Abilities of Man*, p. 19, New York: The Macmillan Company, 1927.

³B. R. BUCKINGHAM, *Research for Teachers*, p. 142.

⁴L. M. TERMAN, in the symposium Intelligence and Its Measurement, *Journal of Educational Psychology*, 12: 128 (1921).

aptitudes."¹ He writes: "It is doubtful whether any unitary faculty of general intelligence exists. We have, instead, a large number of more or less specialized potential aptitudes or intelligences."¹ Not only is Hull's statement contradictory to that of Terman, but it is not in agreement either with that given by Stern.

Ogden, presenting the viewpoint of Gestalt psychology, describes intelligence in terms of articulateness of pattern.² Insight is a significant factor in the establishment of configurations (Gestalten). In his discussion of Köhler's experimental research with apes, Ogden says:

An ape makes many clever errors which give valuable insight into his, and indeed, of all intelligence. . . . A clever error is one in which, although the bridge between a desire and its fulfillment is not adequately bridged, the attempt made indicates the nature and the limits of the animal's insight at the time, and under the conditions of the experiment.

Intelligence might be said to be the capacity for adapting to the environment, from Ogden's statement, with the emphasis to be placed on the way in which the adaptation is made. To this extent, Ogden's interpretation of the Gestalt viewpoint seems to agree in substance with Porteus' definition of "progressive adaptability."

The definitions that have been given and discussed above are typical, and they show clearly an amazing lack of similarity. Each of them has within it something of value, and, without question, some phases of intelligent behavior are described by each. An intelligent individual does think abstractly at times; he does learn; and he has a capacity for acquiring and adjusting and adapting. The behavior of an intelligent individual may be described, in part, in terms of various definitions of intelligence, and yet perhaps no one of them will suffice for a complete description.

Apparent from this diversity of thought is the fact that it is impossible to say that any one concept of intelligence is what psychologists in general think intelligence is. Lacking a commonly accepted definition, it seems worth while to attempt to

¹ C. L. HULL, *Aptitude Testing*, p. 62, Yonkers-on-Hudson, N. Y.: World Book Company, 1928.

² R. M. OGDEN, *Psychology and Education*, pp. 270-271, New York: Harcourt, Brace & Company, 1926.

discover the nature of intelligence by considering the psychological theories as to its nature. It may be possible to delineate or to characterize intelligence, where it is not possible to define it, thus clarifying what may otherwise be an obscure matter.

The Psychological Theories of Intelligence.—Lashley has identified four psychological theories of intelligence that attempt to go beyond mere definition.¹ The first has been called the unit-factor theory, since it regards intelligence as a single variable, as one fundamental function or group of functions, at least in its genetic behavior. Intelligence, according to this theory, is dependent upon anatomical and neurological structures, which are products of heredity, so that intelligence also must be dependent upon heredity. The evidence supporting this theory has been furnished largely by students of heredity, and they obtain support from the data concerning the inheritance of both mental deficiency and mental genius. Lashley writes:

The theory does not rule out special capacities, but suggests that these may be developed to some extent independently of the general capacity which pervades all behavior. There has been no attempt to relate the function of any brain mechanisms. Seemingly the variable might correlate either with the development of a special co-ordinating center, thus harmonizing with doctrines of localization of function, or with the general development of the brain as a whole. The latter in turn might involve complexity of anatomical structure or efficiency of physiological organization, such as delicacy of equilibrium among reaction systems. Assuming the independent development of special capacities, the theory would seem to be more in harmony with doctrines of restriction of intelligence to the activity of association centers than with the belief that the cerebrum acts either as a whole or by summation of the specialized functions of its parts.

The second theory, that advocated by Spearman, has been formulated in great detail. Because of this fact, and also because it probably, with certain modifications developed largely by Kelley and Thurstone, has the support of the majority of psychologists interested in the problem of intelligence, it merits particular attention. According to this theory, the mental ability involved in the doing of any intellectual task or operation may be resolved into two factors. The first of these is a general

¹ K. S. LASHLEY, *Brain Mechanisms and Intelligence*, pp. 1-3, Chicago: The University of Chicago Press, 1929.

factor g common to all mental ability, underlying all cognitive performances of any and every kind. This general factor, a central fund of intellectual energy, varies freely from individual to individual, but it remains constant for any one individual. Similarly, g has been found to have a much greater relative influence, or "weight," in some abilities that have been measured than in others.

When the totality of cognitive operations is thus served by a general factor in common, it is obvious that each different intellectual task or operation or function must be served by some specific elements peculiar to it, and so we discover the existence of the second of the two factors s , which is specific to the given ability necessary in successfully doing the particular intellectual task or operation or to that task or operation and closely similar tasks and operations. When specific factors are common to two or more identifiable different mental tasks or operations but not to all, they are known as group factors. The specific factors involved in any one mental operation or task may be entirely different from those involved in all other such tasks. Specific factors vary from individual to individual and also for any one individual from one ability to another.

It is difficult, without careful analysis, to determine the presence of these two factors and their influence in the doing of any given cognitive act, but means have been found for measuring these things.¹ Intelligence manifests itself in a wide variety of ways, and it is extremely difficult even by careful observation to detect any common characteristic running through a diversity of intellectual operations. The general factor may almost completely saturate one mental operation and be present only slightly in another, and varieties of special factors may be present or absent in two apparently similar mental operations. The nature of the intellectual make-up involved in the doing of any cognitive act can be determined accurately only by proper scientific analysis.

Spearman suggests the hypothesis that g measures something of the nature of an energy. He writes:

Such a general and quantitative factor . . . might be conceived in an infinitude of different ways, including those which would assign to it the most subtle, abstract, or complex constitution. But a readily

¹ SPEARMAN, *The Abilities of Man*, Appendix, pp. i-xxiii.

intelligible hypothesis was suggested to be derivable from physiology. The factor was taken, pending further information, to consist in something in the nature of an "energy" or "power" which serves in common the whole cortex (or possibly, even, the whole nervous system).

But if, thus, the totality of cognitive operations is served by some general factor in common, then each separate operation must necessarily be further served by some specific factor peculiar to it. For this factor also, a physiological substitute has been suggested, namely, the particular group of neurons specially serving the particular kind of operation. These neural groups would thus function as alternative "engines" into which the common supply of "energy" could be alternately distributed.¹

This explanation of the nature of g and s has met with the opposition of students of nerve conduction.

This theory is somewhat similar to the unit-factor theory, in assuming the existence of a general factor plus special abilities. It differs in making more definite assumptions concerning the specific contributions of g and s to given mental abilities.

The viewpoint of the third theory is that intelligence is an algebraic sum of diverse mental capacities. This might be termed the "synthetic theory," since it holds that intelligence consists of distinct elements, variously interrelated. Thorndike has given a clear formulation of this aggregate theory of intelligence. His hypothesis is that quality of intellect depends upon quantity of neural connections. His statement of the hypothesis follows:

Let c represent whatever anatomical and physiological fact corresponds to the possibility of forming one connection or association or bond between an idea or any part or aspect or feature thereof and a sequent idea or movement or any part or aspect or feature thereof. Then if individuals I^1, I^2, I^3, I^4 , etc., differing in the number of c 's which they possess, but alike in other respects, are subjected to identical environments, the amount or degree of intellect which any one of them manifests, and the extent to which he manifests "higher" intellectual processes than the other individuals, will be closely proportional to the number of c 's which he possesses. If we rank them in order by intelligence examination scores, the order will be that of the number of c 's. If we rank intellectual processes in a scale from lower, such as mere information, to higher, such as reasoning, the individuals who manifest the higher processes will have the greatest number of c 's. . . .

¹ C. SPEARMAN, *The Nature of "Intelligence" and the Principles of Cognition*, p. 6, New York: The Macmillan Company, 1927.

Our hypothesis limits itself to the original capacity. If by original nature, apart from all training, a man possesses tendencies to be right rather than wrong in his judgments, to hold true rather than false ideas, to make justifiable rather than unjustifiable inferences, more or less than other men, in so far forth those tendencies are due to his having more or fewer *c*'s than other men. . . .¹

Thorndike later qualifies this statement somewhat by saying that "a capacity for having the neurones act with reference one to another, that is, with integration . . ." ² may also be a factor in the functioning of mental ability. This integrative capacity, he suggests, may be largely irrespective to the number of *c*'s. In addition to this, Thorndike mentions various motivating agents as further conditions for the expression of potential intelligence determined by the number of bonds.

An attempt has been made to show that the theories advocated by Spearman and by Thorndike are not necessarily in disagreement as to the modality of intelligence. Pintner has stated the argument as follows:

There is . . . in Thorndike's view of the problem no need for two factors, general and specific. Yet at the same time Thorndike's theory of intelligence would not exclude the possibility of a general factor such as Spearman has proposed. If it should eventuate that there is a general factor common to all intelligence, then in Thorndike's terminology, this would mean that certain *c*'s are common or necessary to all intellectual tasks. Group factors, rather than general factors, would mean that certain combinations of *c*'s are necessary for certain tasks and other combinations for other tasks.³

The fourth theory may be deduced from the writings of the Gestalt psychologists. According to this, an intelligent act is one in which the elements of a situation are united to form a significant configuration, or whole. Ogden writes: "Acts . . . in which a configuration is defined, and an at first inarticulate mass becomes an articulate pattern, are the prototypes of intelligent behavior."⁴ Koffka believes that a "test of intelligence involves

¹ E. L. THORNDIKE, *The Measurement of Intelligence*, pp. 415-417, New York: Bureau of Publications, Teachers College, Columbia University, 1927.

² *Ibid.*, p. 431.

³ R. PINTNER, *Intelligence Testing-Methods and Results*, p. 71, New York: Henry Holt & Company, 1931.

⁴ OGDEN, *Psychology and Education*, p. 263.

a configural function." He writes: "Strictly speaking, the intelligence of an animal depends upon the configurative character of its receptor processes."¹ Lashley writes concerning this theory:

The theory seems to involve the conception that each increment of intelligence involves a quantitatively different organization from the last. This would outlaw the current conceptions of quantitative differences in intelligence and the attempts to express the growth of intelligence upon a continuous scale. It implies also a distinct type of neurological theory, departing widely from the more generally accepted views of the fundamental nervous mechanisms and precluding any simple quantitative relationship between neural organization and intellectual level.

This theory is opposed to Thorndike's that intelligence is conditioned by the number of neural connections that one has by original nature. The criticism may be clarified by another quotation:

Attention should be especially directed to the futility of attempting to derive intelligence and the higher mental faculties in general from reflexes, habits, or any form of fixed or determinate behavior. On the contrary, these owe their origin to the more labile and plastic components of behavior, which are determined, if you like, though not by rigid innate organization, but rather by individual experience acting through and upon the innate units and recombining these in new patterns. The nervous system is more than an aggregate of reflex arcs and life is more than reaction to stimuli.²

Although this theory has not been elaborated in relation to the problem of individual differences, it implies that individual differences in intelligence result from the possession of fewer or more modes of relating the elements in intellectual experience. Just how mental ability functions in relating the diverse elements in a situation into an articulate and meaningful configuration is not known.

Considered thus far in this chapter have been various definitions of the term intelligence and certain psychological theories concerning the nature of intelligence. Disagreement is found,

¹ K. KOFFKA, *The Growth of the Mind*, pp. 232, 251, New York: Harcourt Brace & Company, 1925.

² C. J. HERRICK, *Neurological Foundations of Animal Behavior*, p. 234, New York: Henry Holt & Company, 1924.

both in the definitions and in the theories, so that it is not possible to say definitely with any certainty just what intelligence is. Yet what has been presented aids in clarifying thinking on the subject. Something is gained in a discussion of the nature of intelligence by a consideration of its characteristics or attributes. It is in this direction that we must now turn for aid in understanding intelligence and its functioning.

Characteristics of Intelligence.—Thorndike has distinguished four attributes of intelligence¹: (1) level, (2) range, (3) area, and (4) speed.

Level refers to the degree of difficulty present in specific intellectual tasks, *i. e.*, those tasks which judgment dictates require the successful functioning of intelligence in their achievement. Difficulty may be defined as the characteristic involved in the given task that causes persons to respond to it incorrectly. A difficult intellectual task is one the successful accomplishment of which is arduous and attended with and occasioning trouble during the solution. Thorndike writes: "Intellectual difficulty . . . [is] . . . that characteristic of a task, an increase in which reduces the number of intellects who can succeed with it, eliminating those esteemed unintelligent more rapidly than those esteemed intelligent."² Elsewhere he comments that "Psychologists and scientists and sensible people in general readily rank tasks as easy or hard for intellect and would accept the principle that 'Other things being equal, the harder the tasks a person can master, the greater is his intelligence.'"³

Range of intelligence refers to breadth of mental ability. It may be thought of as supplementing difficulty in describing intelligence, since it is determined by the number of tasks at any level of intellectual difficulty which an individual can solve or achieve correctly. If all such tasks and situations within the experience of man could be gaged accurately with respect to the amount of intelligence required to solve or to achieve them, then it would be possible to group these tasks at each difficulty level for intelligence, and the number of such tasks that a given individual could solve at any given level of difficulty would represent the range of intelligence for that person at that level. Unfortunately, it is

¹ THORNDIKE, *The Measurement of Intelligence*, pp. 22-23.

² *Ibid.*, p. 59.

³ *Ibid.*, p. 22.

not possible to determine accurately the difficulty of all intellectual tasks, so that in practice it is not possible to determine accurately the exact range of the intelligence of any given individual. It is possible, of course, to estimate range of intelligence from a sampling of the types of tasks at any given difficulty level that an individual can solve correctly. In theory, if an individual can solve one intellectual task at a given difficulty level, he should be able to solve all other intellectual tasks at the same level. Practically this is impossible, since all individuals have not had equal opportunities for intellectual experience, and a lack of knowledge of facts at lower levels or a lack of skill or knowledge at the given difficulty level would prevent the correct solution of tasks at levels where the individual might be successful with still other tasks. Range, then, like difficulty does not permit a complete and correct inventory of the intelligence of any given individual.

Area may be thought of as the product of intellectual level and intellectual range. For convenience, it may be thought of as the total number of tasks of an intellectual nature in the correct solution of which an intellect is successful, *i.e.*, the summation of all the ranges of intelligence at all the levels possessed by an individual. It has been pointed out that it is not possible in practice to determine range of intelligence accurately. Because of this fact, area of intelligence must be thought of as not possessing great practical significance, however significant it may be theoretically.

Speed is the rapidity with which individuals respond to situations, to which must be added, when intellect is under consideration, the swiftness or celerity with which they solve intellectual tasks correctly. Speed is measured in terms of the number of tasks correctly completed and in terms of the amounts of time required for these achievements. Thorndike writes: "The units of number and time are indeed so much more convenient and intelligible than units of difficulty that there is a strong natural tendency in those who devise instruments for measuring intellect to let their measurements depend largely upon the number of correct responses and the speed of producing them. . . ." ¹ Such a procedure introduces errors, since both level and range must be considered in obtaining any measurement of intelligence.

¹ THORNDIKE, *The Measurement of Intelligence*, pp. 32-33.

However, speed in intellectual production is highly desirable, and there can be no doubt that an intellect that functions with celerity and dispatch in accomplishing successfully a given group of intellectual tasks correctly is better than one that functions more sluggishly but with equal accuracy.

The preceding paragraphs have attempted to delineate the attributes of intellect, so that the factors that enter into successful intellectual function may be known, even though we do not know exactly the nature of general intelligence itself. The products of intelligence are more or less easily recognized, but intelligence itself is obscure. It is possible to think of different kinds or types of intelligence, and it is with these that we shall now be concerned.

Kinds of Intelligence.—One classification of the types of intelligence includes (1) abstract or academic intelligence, (2) social intelligence, (3) concrete (including mechanical) intelligence, and (4) aesthetic intelligence. Just what each of these draws from or contributes to general intelligence is not known, but each of the types is readily recognized in and of itself.

Abstract intelligence is concerned with general concepts and with ideas. It is demonstrated when an individual responds adequately to words, definitions, and symbols. This type of mental ability is essential for success in the study of school subjects; it is essential to learning and to scholarly achievement.

Social intelligence is demonstrated by individuals through living successfully in a social environment. Adequate and successful living demands that people engage in friendly companionship, in political and economic cooperation and intercourse, and in desirable social relations. The person with high social intelligence receives social approbation much more readily than does the person with low social intelligence. For this reason, the person with high social intelligence "gets the votes or the business" rather than the person with low social intelligence.

Concrete intelligence requires individuals to respond adequately to particular things or instances, to realities and to actual items. This kind of intelligence varies from mere manipulation of objects or mechanical aptitude on the one hand to the ingenuity of the inventor of complex mechanical apparatus on the other.

The nature of aesthetic intelligence is not yet clearly established, and no measures of aesthetic appreciation have yet been satisfactorily validated. Aesthetics may be defined as the science

of the beautiful.¹ It has been said that aesthetics is concerned primarily with the emotions aroused by beauty.² Controversially, there is the attitude that aesthetic appreciation is concerned with objects, with things or ideas, and that real emotion is not generated by other than real stimuli. Taking a middle position in writing of literary appreciation, Judd says:

Literary appreciation is the product of intellectual understanding and also of active response on the part of the reader. Where there is no ability to understand, there is a lack of appropriate response; where the response is absent or inappropriate, the resultant emotion will be deficient or perverted. . . .

. . . [The reader] must cultivate a personal response. . . . This personal response requires an intellectual understanding of the meaning of the selection. In addition to the intellectual understanding, there must be an emotional reaction.³

TABLE XIX.—MEASURES OF RELATIONSHIP BETWEEN ABSTRACT AND SOCIAL INTELLIGENCE (AS MEASURED BY THE MOSS ET AL., SOCIAL INTELLIGENCE TEST)

Author	Reference No. ¹	Test of abstract mental ability	Year	N	r	P.E.
Grosvenor.....	5, p. 656	Terman	1927	136	0.75	0.03
Broom.....	2, p. 428	Thorndike	1928	258	0.56	0.03
Garrett and Kellogg..	4, p. 125	Thorndike	1928	118	0.42	0.05
Hunt.....	6, p. 327	George Washington University Mental Alertness Test	1928	243	0.54	0.03
		O'Rourke College Entrance Test		102	0.57	
		Thorndike		118	0.42	0.05
		Brown University Test		96	0.56	
		McCall Multi-mental		130	0.25	
Pintner and Upshall..	13, p. 370	Thorndike and CAVD Verbal Intelligence Test	1928	33	0.68	0.06
Broom.....	3, p. 404	Thorndike	1930	646	0.60	0.02
Scudder and Raubenhaimer.....	14, p. 122	Terman	1930	107	0.48	0.05
Strang.....	18, p. 270	T. C. Verbal Intelligence Test				
		Thorndike and CAVD	1930	311	0.44	0.04
Stagner.....	15, p. 456	American Council Psychological Examination	1932	62	0.46	0.07

¹ References are listed in special bibliography at the end of the chapter.

² *New Standard Dictionary of the English Language*, pp. 139, 854, New York: Funk & Wagnalls Company, 1929.

³ *Encyclopaedia Britannica* (9th ed.), Vol. I, p. 212, New York: Encyclopaedia Britannica, Inc., 1875.

⁴ C. H. JUDD, *Psychology of Secondary Education*, p. 219, Boston: Ginn and Company, 1927.

270 MEASUREMENTS IN THE ELEMENTARY SCHOOL

TABLE XX.—RELATIONSHIP BETWEEN MECHANICAL AND ABSTRACT (ACADEMIC) INTELLIGENCE

Author	Reference number ¹	Test of abstract mental ability	Test of mechanical mental ability	Year	N	r	P.E.
Stenquist	17, p. 7	Average of 6 tests, academic intelligence: National Intelligence Tests A & B, Haggerty Intelligence Examination Δ, Otis Intelligence Test, Myers Mental Measure, Thorndike Visual Vocabulary Scale.	Average of 4 tests, Stenquist tests: Assembly Series, I & II Picture (group) Tests, I & II	1922	275	0.21	0.04
MacQuarrie	11, p. 329-337	Group mental test	MacQuarrie	1927		Less than 0.20; one r as low as 0.002, all positive.	
Stein	16, p. 391	Kohs Army Alpha	Stenquist I Stenquist II Stenquist II MacQuarrie	1927	15 15 15 15	-0.12 -0.45 0.10 -0.35	0.18 0.14 0.18 0.16
Isaac	7, p. 36	National Intelligence Tests	Stenquist	1928	56 16	0.25 0.07	0.08 0.17
Peterson and Lanier	12, p. 142	Otis Self-administering Test	Minnesota Tapping I Minnesota Speed of Movement Minnesota Substitution Minnesota Cancellation Minnesota Form Board Minnesota Tapping II Stenquist I Stenquist II	1929	130 130 130 130 130 60 53	-0.14 -0.10 0.22 -0.01 0.33 -0.05 0.07 0.09	0.06 0.06 0.05 0.06 0.06 0.09 0.09
Scudder and Raubenhimer	14, p. 122	Terman	MacQuarrie Stenquist I O'Rourke Stenquist Assembly, II	1930	105 102 79 43	0.06 0.10 0.26 0.19	0.07 0.07 0.07 0.10
Attenborough and Farber	1, pp. 140-161	Stanford Binet Otis Group Test Primary A Healy Picture Completion II Knox Cube Imitation Porteus Maze Gaw Cube Construction Dearborn Formboard	Stenquist III Stenquist III	1934	80	0.45	0.06 0.05 0.07 0.07 0.40 0.07 0.07
Jorgensen	8, p. 98	Army Alpha	MacQuarrie	1934	172	0.24	0.05
Weinberg and Broom	Unpublished	Thorndike	MacQuarrie	1934	55	0.18	0.07

¹ References are listed in a special bibliography at the end of the chapter.

It is clear that the nature of aesthetic appreciation is yet to be finally established. It may be that it is made up of what has been defined as abstract intelligence plus emotion. Because of this uncertainty, and because it has been impossible up to the present time to measure aesthetic appreciation with certainty, no further discussion will be given here on the subject.

The measures of relationship involving findings yielded by tests of academic intelligence, of social intelligence, and of mechanical (low-level concrete) intelligence have been assembled from research data in Tables XIX to XXI inclusive. The findings in Table XIX indicate that the relationship between academic mental ability and social intelligence is marked. It has been shown elsewhere, however, that this marked agreement is largely due to the self-correlation of silent reading comprehension, a common variable measured by both types of tests.¹ The true situation is low and positive agreement between the variables. If an individual, then, has high abstract or academic intelligence, he may or may not have high social intelligence; the chances are almost (but not quite) as good that he will have low as that he will have high social intelligence.

TABLE XXI.—MEASURES OF RELATIONSHIP BETWEEN MECHANICAL APTITUDE AND SOCIAL INTELLIGENCE (AS MEASURED BY THE MOSS ET AL., SOCIAL INTELLIGENCE TEST)

Author	Reference No. ¹	Test of mechanical mental ability	Year	N	r	P.E.
Hunt.....	6, pp. 327, 328	O'Rourke	1928	130	0.22	0.06
		MacQuarrie		126	0.11	0.06
Scudder and Raubenhaimer.	14, p. 122	MacQuarrie	1930	107	0.06	0.06
		Stenquist I		103	0.00	0.07
		O'Rourke		79	0.34	0.07
		Stenquist Assembly II		45	0.03	0.10
Broom.....	Unpublished	MacQuarrie	1931	57	0.29	0.08

¹ References are listed in a special bibliography at the end of the chapter.

The findings in Table XX indicate that the relationship between abstract mental ability and mechanical aptitude is low

¹ M. E. BROOM, A Further Study of the Validity of a Test of Social Intelligence, *Journal of Educational Research*, 22: 403-405 (1930); 23: 54 (1931).

and positive; *i.e.*, there is only some slight agreement above chance agreement between these two variables. The facts should be interpreted as were those for Table XIX. The data in Table XXI indicate that mechanical aptitude and social intelligence, as measured by the available tests, also correlate to a low and positive degree. The interpretation is similar again to that for Table XIX. These kinds of intelligence are obviously distinct. High degrees of ability in any one of the three fields may or may not be accompanied by similar high degrees of good ability in the other types of intellect.

Summary.—The discussion in this chapter has attempted to furnish the student with information concerning the origins of man's thinking about intelligence; man's attempt to define and to describe intelligence, including the psychological theories as to the nature and characteristics of intelligence; and the recognized kinds or types of intelligence. No effort has been made to present research findings, except in the latter section of the chapter. The effort has been rather to bring to the student in simple terms the thinking of many psychologists. It is obvious that progress has been made by psychology in its thinking about intelligence, and it is equally obvious that the problem has not yet been conquered, since man does not yet know with certainty what intelligence is or exactly how it functions.

References for Tables XIX, XX, and XXI

1. J. ATTENBOROUGH and M. FARBER, The Relation between Intelligence, Mechanical Ability, and Manual Dexterity, *British Journal of Educational Psychology*, 4: 140-161 (June, 1934).
2. M. E. BROOM, A Note on the Validity of a Test of Social Intelligence, *Journal of Applied Psychology*, 12: 426-428 (1928).
3. M. E. BROOM, A Further Study of the Validity of a Test of Social Intelligence, *Journal of Educational Research*, 22: 403-405 (1930); 23: 54 (1931).
4. H. E. GARRETT and W. N. KELLOGG, The Relation of Physical Constitution to General Intelligence, *Journal of Experimental Psychology*, 11: 113-129 (1928).
5. EDITH L. GROSVENOR, A Study of the Social Intelligence of High School Pupils, *American Physical Educational Review*, 32: 649-657 (1927).
6. THELMA HUNT, Measurement of Social Intelligence, *Journal of Applied Psychology*, 12: 317-344 (1928).
7. RACHEL ISSACS, A Comparison of the National Intelligence Test Results and the Stenquist Test Results, in J. C. Bell and A. L. Suhrie, *Contributions to Education*, Vol. II (New York Society for the Experimental

- Study of Education). Yonkers-on-Hudson, N. Y.: World Book Company, 1928.
8. C. JORGENSEN, Analysis of Some Psychological Tests by the Spearman Factor Method, *British Journal of Educational Psychology*, 14 (Part I): 98 (February, 1934).
 9. VIVIENNE R. MACLATCHY, A Theoretical and Statistical Critique of the Concept of Social Intelligence, *Journal of Abnormal and Social Psychology*, 24: 217-220 (1929).
 10. T. W. MACQUARRIE, *A Measure of Mechanical Ability*. Los Angeles: Associated Students Store, University of Southern California, 1925.
 11. T. W. MACQUARRIE, A Mechanical Ability Test, *Journal of Personnel Research*, 5: 329-337 (January, 1927).
 12. J. PETERSON and L. H. LANIER, Comparative Abilities of Negroes and Whites. *Mental Measurement Monographs* 5. Baltimore: Williams & Wilkins Company, 1929.
 13. R. PINTNER and C. C. UPSHALL, Some Results of Social Intelligence Tests, *School and Society*, 27: 369-370 (1928).
 14. C. R. SCUDDER and A. S. RAUBENHEIMER, Are Standardized Mechanical Aptitude Tests Valid? *Journal of Juvenile Research*, 14: 120-123 (1930).
 15. ROSS STAGNER, The Intercorrelation of Some Standardized Personality Tests, *Journal of Applied Psychology*, 16: 453-464 (October, 1932).
 16. M. L. STEIN, A Trial with Criteria of the MacQuarrie Test of Mechanical Ability, *Journal of Applied Psychology*, 11: 391-393 (1927).
 17. J. L. STENQUIST, *Manual of Directions for the Stenquist Mechanical Aptitude Tests*. Yonkers-on-Hudson, N. Y.: World Book Company, 1922.
 18. RUTH STRANG, Relation of Social Intelligence to Certain Other Factors, *School and Society*, 32: 268-272 (1930).

Readings for Further Study

1. W. F. DEARBORN, *Intelligence Tests*. Chaps. III, IV. Boston: Houghton Mifflin Company, 1928.
2. N. FENTON and D. A. WORCESTER, *Educational Measurements*. Chap. II.
3. J. M. FLETCHER, *Psychology in Education*. Pp. 97-99, 108-116. Garden City, New York: Doubleday, Doran & Company, Inc., 1934.
4. F. N. FREEMAN, *Mental Tests*. Boston: Houghton Mifflin Company, 1926. Read pertinent chapters.
5. H. E. GARRETT, *Great Experiments in Psychology*. Pp. 3-49. New York: D. Appleton-Century Company, Inc., 1930.
6. C. R. GRIFFITH, *An Introduction to Educational Psychology*. Pp. 533-564. New York: Farrar & Rinehart, Inc., 1935.
7. T. L. KELLEY, *Crossroads in the Mind of Man*. Palo Alto, Calif.: Stanford University Press, 1928.
8. K. S. LASHLEY, *Brain Mechanisms and Intelligence*.
9. G. MURPHY, *Historical Introduction to Modern Psychology*. Pp. 347-372. New York: Harcourt, Brace & Company, 1929.

10. J. PETERSON, *Early Conceptions and Tests of Intelligence*.
11. R. PINTNER, *Intelligence Testing—Methods and Results*. Read pertinent chapters.
12. R. PINTNER, The Individual in School: General Ability, in C. Murchison, ed., *The Foundations of Experimental Psychology*. Pp. 685–701. Worcester, Mass.: Clark University Press, 1929.
13. G. C. SCHWESINGER, *Heredity and Environment*. Pp. 5–85. New York: The Macmillan Company, 1933.
14. C. SPEARMAN, *The Abilities of Man*.
15. E. L. THORNDIKE, *The Measurement of Intelligence*.
16. E. W. TIEGS, *Tests and Measurements for Teachers*. Chap. XVIII.
17. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Pp. 317–330.
18. K. YOUNG, The History of Mental Testing, *Pedagogical Seminary*, 31: 1–48 (1924).

Questions and Exercises

1. Formulate your own definition of intelligence, and show in what ways it is similar to or different from at least five other definitions of intelligence.
2. Discuss the characteristics of intelligence. Illustrate each, using materials from ordinary experience.
3. If there are several kinds of intelligence, such as abstract, mechanical, and social, how may each be defined? To what extent do these different kinds of intelligence overlap?
4. List a number of mental characteristics or traits that in your opinion differentiate persons of superior intelligence from persons of retarded intelligence. Comment on your reasons for each selection.
5. In what way or ways has the use of mental tests contributed to or retarded the development of man's knowledge of what intelligence really is?
6. Make a bibliography of the Thorndike-Spearman controversy to the present time. Summarize briefly the contentions of both schools of thought, and state the present situation with respect to this controversy.
7. What physiological and anatomical facts have been considered in connection with the formulation of the different psychological theories as to the nature of intelligence?

CHAPTER XII

INTELLIGENCE AND ITS MEASUREMENT: MENTAL TESTS

The preceding chapter has presented information briefly summarizing the history of mental measurement together with statements giving the present thoughts of psychologists and educators as to the nature of intelligence. This information is necessary for an understanding of mental measurement. It tells little, however, concerning the nature of mental tests or of their uses in practical classroom situations. The purpose of this chapter is to present the theory of mental test building and to supply information concerning the content of mental tests and data relating to mental tests. In Chapter XIII will be given the uses of mental tests in practical classroom situations.

A Philosophy of Mental Measurement.—The existence of intelligence and evidence of intelligence have intrigued man for many centuries. Men assume the existence of intelligence in all human beings and assume, furthermore, that it can be measured by means of tests. In Chapter XI, it has been shown that different types or phases of intelligence exist. No general tests of intelligence have yet been devised. The measuring instruments that now exist measure one of the several types or phases of intelligence that have now been identified. The great majority of intelligence tests now available for use in elementary-school grades measure academic or abstract intelligence. This fact is demonstrated by the data given in Table XXII.

The information in Table XXII indicates the existence of a marked to high degree of relationship between total academic achievement, as measured by batteries of school achievement tests, and mental ability, as measured by so-called intelligence tests. It is perhaps worth noting that similar findings occur in widely scattered geographical areas and in different school grades as well as with different mental and achievement tests. The similarity of the findings, obtained under different conditions at different times and in different geographical areas, indicates that

TABLE XXII.—MEASURES OF RELATIONSHIP BETWEEN TOTAL SCORES (INCLUDING THEIR DERIVATIVES) YIELDED BY CERTAIN MENTAL TESTS AND BY CERTAIN BATTERIES OF SCHOOL ACHIEVEMENT TESTS¹

Author	Year	Mental test	Battery of school achievement tests	School grades	N	$r \pm P.E.$	Location
B. H. Broom.....	1929-1933	Otis Classification	Test, I vs. II	7	3473	0.632 \pm 0.007	El Paso, Tex.
		Otis I.Q.	Otis E.Q.	7	3473	0.788 \pm 0.005	El Paso, Tex.
M. E. Broom.....	1929	McCall	New Stanford	4 to 6	273	0.732 \pm 0.019	Chula Vista, Calif.
	1929-1930	McCall	New Stanford	4 to 6	324	0.752 \pm 0.016	Chula Vista, Calif.
	1929	McCall	Public School Achievement	4 to 8	575	0.805 \pm 0.010	El Centro, Calif.
DeSilva.....	1933	Terman	New Stanford	7	261	0.734 \pm 0.019	San Diego, Calif.
Jordan ²	1930	IER, Test 3	Stanford	4 to 7	271	0.820 \pm 0.013	North Carolina
		IER, Test 3	Stanford	4 to 7	176	0.739 \pm 0.023	North Carolina
Perry.....	1933	Thurstone	Iowa HS Content	13	178	0.749 \pm 0.022	Long Beach, Calif.
		Thurstone	Iowa HS Content	13	144	0.765 \pm 0.023	Long Beach, Calif.
Skiff ³	1938	Henmon-Nelson	Sones-Harry	10	100	0.628 \pm 0.041	El Paso, Tex.
Witty and Taylor.....	1928	Stanford-Binet	Stanford	4 to 6	522	0.634 \pm 0.017	Lawrence, Kans.
		McCall	Stanford	4 to 6	522	0.780 \pm 0.012	Lawrence, Kans.
Wright.....	1931	Terman	Sones-Harry	9	81	0.702 \pm 0.038	San Pedro, Calif.

¹ The studies from which these data have been taken have been listed in a separate bibliography at the end of this chapter. The major part of this information originally was assembled in the following discussion: B. H. Broom, A Note on the Relationship between Mental Ability and School Achievement, *El Paso Schools Standard*, 14 (No. 1): 14-17 (October, 1936).

² Certain of these facts were given by Dr. Jordan in a letter dated Apr. 24, 1934.

³ From an unpublished study by Josephine M. Skiff.

this relationship is one that might be anticipated generally. Whatever is measured by mental tests is also measured by batteries of school achievement tests. Since batteries of school achievement tests definitely measure academic achievement, it follows that to a very marked degree mental tests measure academic intelligence.

The difference in the purpose of these types of tests is possibly worth noting. The battery of school achievement tests attempts to evaluate what has happened already, to measure what pupils have learned in school classrooms in a variety of subjects. The mental test attempts to evaluate pupils' ability to learn, to provide a means for predicting at some future time the degree of success that pupils may attain in academic work.¹

Types of Mental Tests.—There are many mental tests available for use in schools, and they cover the entire range of school grades from the nursery school to the graduate schools of universities. Naturally, the grade coverage of any one test is more or less limited, although some tests are advertised as covering a range of six to eight school grades. In addition, there are mental tests devised for use with infants.

These tests fall into two general types, individual and group. Individual mental tests include all those which were devised primarily for administration to one person at a time. Group tests, on the other hand, can be given to from two to several hundred persons at one time. The giving of individual mental tests requires much more training than can be obtained in a single course in educational measurements; it is a job for experts. For this reason, individual mental tests are not treated extensively in this discussion.² Group tests can be given properly by any classroom teacher willing to follow exactly the directions

¹ The predictive functions and values of mental tests will be discussed later in this chapter and in Chap. XIII.

² The student who is interested in obtaining information about individual mental tests can read with profit the following books:

A. F. BRONNER, W. HEALY, G. M. LOWE, and M. E. SHIMBERG, *A Manual of Individual Mental Tests and Testing*, Boston: Little, Brown & Company, 1928.

F. N. FREEMAN, *Mental Tests*, Chaps. IV-VI.

H. E. GARRETT and M. R. SCHNECK, *Psychological Tests, Methods, and Results*, Part II, Chaps. I, II, New York: Harper & Brothers, 1933.

S. C. KOHS, *Intelligence Measurement*, New York: The Macmillan Com-

furnished with the several group mental tests. The discussion in the remainder of this chapter will be concerned with group mental tests.

The Construction of Group Tests.—What a person has learned prior to any given life age, whether in schools or elsewhere, may be taken as an index of his intelligence. Certain fundamental assumptions must be accepted in this connection: (1) What one has learned can be measured; (2) all persons measured have had equal opportunities to learn the materials included in the test; (3) all persons tested have had equal inclination to learn the materials of the test. Obviously, these assumptions are only partially true.

The task of the test maker is first to determine what people have learned during certain periods of life. Then, still observing human behavior in general, the psychologist engaged in devising a mental test attempts to select certain types of situations in which learning takes place and that demand the exercise of comprehension, memory, ingenuity, alertness and speed of response, and the like—what men call the exercise of intelligence. These situations must be within the capacity and the experience of the individuals to be tested.

It is then necessary for the test maker to select materials that can be employed within the typical learning situations. These form the subject matter of the mental test. They must be such that they can be used quickly and with little or no apparatus to secure the functioning of the mental capacities that it is desired to measure. The materials may be printed passages to be read and comprehended, lists of words to be defined, mathematical problems to be solved, and so on. The mental test is so devised that the operations through which the person responding to the test must go to achieve success are those which intelligent people in general complete in achieving success in similar problem situations.

What is essential, of course, in a mental test is the presence of a considerable variety of different types of exercises in the instru-

pany, 1923.

R. PINTNER, *Intelligence Testing—Methods and Results*, Chap. VI.

S. D. PORTEUS, *The Maze Test and Mental Differences*, Vineland, N. J.: The Training School, 1933.

L. M. TERMAN and M. A. MERRILL, *Measuring Intelligence*, Boston. Houghton Mifflin Company, 1937.

ment. Mental ability is seldom measured accurately and completely by success in one type of problem situation only, for it is seldom highly specialized. To evaluate mental ability properly, therefore, it is necessary to tap numerous mental capacities. The entire test must be comprehensive yet flexible enough to involve the use of many different kinds of mental operations, if the persons tested are to reveal their mental ability adequately and completely.

Proper attention must be given also to arranging the test materials according to their difficulty as well as according to their type. If this is not done, persons responding to the test will spend too much time on very difficult items occurring early in the test and so fail to complete easier items which occur later in the instrument. The content of each type must include items easy enough to permit some success by all persons responding to the test; and also there must be included content so difficult that few if any responding to the test can do all of the tests in the entire instrument successfully in the allotted time. Intervening exercises must be arranged in order, from easy to difficult. If this is not done, it cannot be said that the actual mental capacity of any person responding to the test has been measured. This is particularly true of anyone who does successfully either none or all of the tasks included in the instrument.

It was indicated above that all persons tested must have had equal opportunity and equal inclination for learning the materials of the test if the instrument is to yield valid indices of mental ability. Obviously, individual differences with respect to different types of test content will exist. Yet the final result yielded by the test represents a kind of average of achievement on the several different kinds of tasks of different types. One individual may be able to solve difficult exercises of one type and yet have little success with much easier exercises of another type. Since mental tests normally include several different kinds of content, with each type arranged according to difficulty, individual differences with respect to the different kinds of content are likely to "average out," with the final result indicating mental ability better than could any test with but one type of content. In other words, the fundamental assumptions apply more to the total test, with its variety of types and difficulty of content, than to any individual type or difficulty of content included in the instrument.

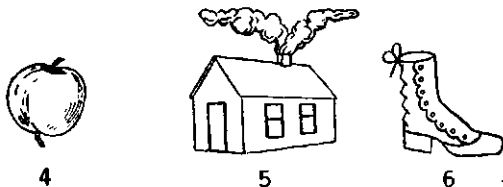
It is now proper to consider the specific types of content included in group tests devised for use in elementary-school grades.

Specific Types of Mental-test Content.—With few exceptions, group tests require the use of English, either in comprehension of the directions for doing the test or in the test content proper. This statement seems necessary, since the following illustrative types of items in some instances would seem to indicate that the tests are in part nonverbal. It should be emphasized that these few items do not indicate the character of the complete tests from which they are chosen and that portions of tests, quoted below and in the following discussion, do not indicate either what the complete tests measure or the manner in which they are administered, since complete directions are not given with the illustrative materials.¹ The illustrations of specific types of mental-test content follow.

1. *Aesthetic differences.* (From the Pintner-Cunningham Primary Mental Test.)²



2. *Correction of pictorial absurdities.* (From the Kuhlmann-Anderson Intelligence Tests, Grade I, First Semester.)³

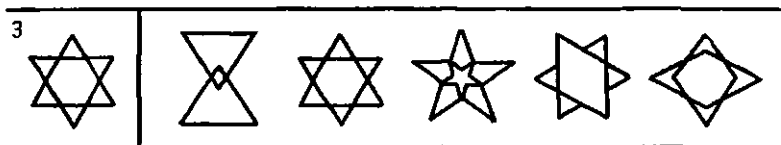


¹ This precaution is taken to prevent the tests from becoming familiar material to persons who might be given them later. When those responding to a mental test are familiar with the instrument, the scores probably will be unreliable.

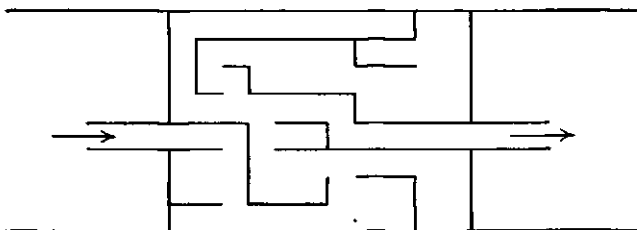
² Published by The World Book Company, Yonkers-on-Hudson, N. Y.

³ Published by the Educational Test Bureau, Minneapolis, Minn.

3. *Identification of similar objects and following directions.*
(From the Detroit Beginning First-grade Intelligence Test, rev. ed.)¹



4. *The maze.* (From the Cole-Vincent Group Intelligence Test for School Entrants.)²



5. *Analogies.* (From the Otis Quick-scoring Mental Ability Tests, Beta Test.)¹

55. A wire is to electricity as (?) is to gas.

1 a flame 2 a spark 3 hot 4 a pipe 5 a stove

6. *Word meaning* (and identification of the word that is different from the remaining four words). (From the Detroit Alpha Intelligence Test.)³

1	1 apple	2 banana	3 lemon	4 pear	5 turkey
2	1 East	2 equator	3 North	4 South	5 West

7. *Arithmetical problems.* (From the California Short-form of Mental Maturity Test, Intermediate S-form.)⁴

1. If a freight train travels at the rate of 20 miles an hour, how many miles will it travel in 4 hours?

Ans.: ^a 5 ^b 20 ^c 24 ^d 60 _____ ¹

¹ Published by The World Book Company, Yonkers-on-Hudson, N.Y.

² Published by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kan.

³ Published by the Public School Publishing Company, Bloomington, Ill.

⁴ Published by the California Test Bureau, Los Angeles.

8. *Number series.* (From the Terman Group Test of Mental Ability.)¹

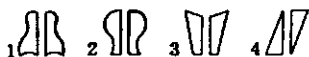
1st Row	8	7	6	5	4	3
2d Row	3	8	13	18	23	28

Two quite different methods have been followed by test makers in presenting the content of mental tests. Probably the more common one might be called the subtest arrangement, in which similar items have been grouped together in definite divisions. An example of this type of grouping is found in the California Short-form Test of Mental Ability, Intermediate S-form, in which the Non-language Tests are divided into three groups: (1) sensing right and left, (2) manipulation of areas, and (3) similarities; the Language Tests also are grouped into three groups: (1) numerical quantity, (2) inference, and (3) vocabulary.

The second and less common type of arrangement has been called the omnibus-cycle arrangement. This is well illustrated by the Otis Quick-scoring Mental Ability Tests, Beta Test. As will be noted in the following excerpt from the test, the items are grouped according to difficulty without regard to type, except that no two adjacent items are similar.

14. At 3 cents each, how many pencils can be bought for 27 cents?

1 27 2 81 3 3 4 9 5 30



15. Three of the four designs at the left are alike.

Which one is not like the other three?

16. Which is the most important reason that words in the dictionary are arranged alphabetically?

1 It is the simplest arrangement. 2 It puts the shortest words first. 3 It enables us to find any word quickly. 4 It is merely a custom. 5 It makes the printing easier.

17. Which one of the five things below is most like these three: a saw, a hammer, and a file?

1 a bottle 2 a pen 3 a screw driver 4 a fork 5 a carpenter..

18. If the following words were arranged in order, which word would be in the middle?

1 luncheon 2 dress 3 undress 4 supper 5 breakfast.....

¹ Published by the World Book Company, Yonkers-on-Hudson, N. Y.

It should be noted, perhaps, that it is customary to time each division of the test separately, when the test content is organized in subtests; whereas with the omnibus-cycle arrangement, a single time limit usually is given for the entire test. This contributes to ease of administration of the latter type of test.

The preceding discussion has dealt somewhat generally with the characteristics of mental tests devised for use in elementary grades. It is now apropos to deal somewhat in detail with certain selected mental tests.

Tests Devised for Use in Elementary Grades.¹—*The Pintner Non-language Primary Mental Test.*²—Frequently, a need arises for a non-language test of mental ability for use with kindergarten and primary pupils, particularly when some members of a group at these grade levels come from homes in which English is seldom or never spoken. Such a test is valuable, too, in equalizing the differences in language opportunities found among children from good and from poor homes.

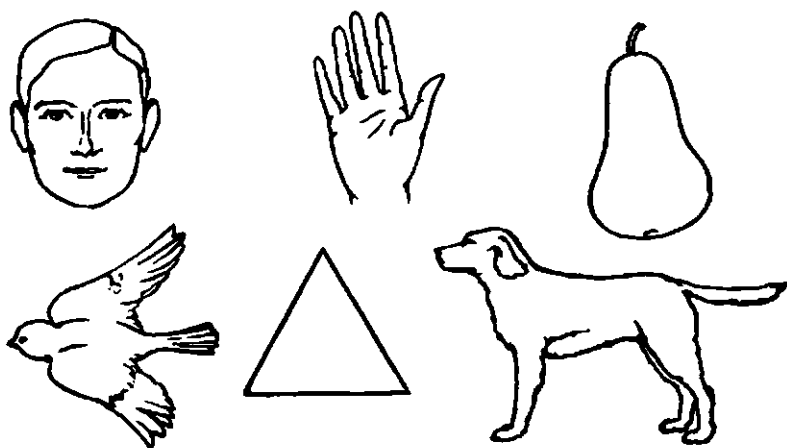
The Pintner Non-language Primary Mental Test is devised for young children in the kindergarten and first two elementary grades. It does not require a knowledge of the English language, nor does it presuppose a knowledge of reading and writing. The directions are given entirely in pantomime in the form of a game. The children responding to the test are not required to give any verbal responses. Ordinarily, 10 children may be tested at once, if they are entering school for the first time. Larger numbers of pupils who are familiar with class instruction can be tested simultaneously.

The type of directions for giving the test and the content can be illustrated readily by quoting from the *Manual of Directions* accompanying the test (page 4) and from the test proper. The procedure and content for page 3 of the test is as follows:

Page 3. Attract attention. Hold up right hand with palm toward children. Point to your palm and then to booklet (not to palm in booklet). Make marking gesture in air. Nod. Hold palm up for 15 seconds. Turn to page 4.

¹ Many tests of mental ability are available for use in elementary-school grades. Those chosen are more or less typical. They have been selected to demonstrate the contents, directions for giving, and data as to the validity and reliability that are characteristic of such tests in general.

² Devised by R. Pintner and published by the Bureau of Publications, Teachers College, Columbia University, New York.



Complete instructions are available for scoring the test. Mental-age equivalents for scores earned on the test are available for years four through nine, and norms are given for the kindergarten and for grades 1 and 2. The test author gives reliability coefficients of 0.90 (111 cases) for grade 1 and 0.86 (129 cases) for grade 2. The author reports also several validity coefficients. These give some indication of how closely this test resembles other tests of mental ability very different in type from this. The coefficients are:

1. With mental age, Stanford-Binet, $r = 0.61$ (80 kindergarten children); $r = 0.78$ for 28 cases, grade level not indicated.
2. With mental age on the Pintner-Cunningham Primary Mental Test, $r = 0.51$ (154 children in kindergarten to grade 2).¹

*The Pintner-Cunningham Primary Mental Test.*²—This test is designed to measure the mental ability of young children just beginning their school career. The content of the test includes pictorial material, and the directions are given verbally. This tends to allow equal opportunity for success to all children, since children in the kindergarten and first grade frequently have varying backgrounds, and few if any of them have sufficient

¹ The information in the preceding paragraph was taken from R. Pintner, *Manual of Directions for the Non-language Primary Mental Test*, pp. 10, 11.

² Devised by R. Pintner and Bess V. Cunningham and published by the World Book Company, Yonkers-on-Hudson, N. Y.

familiarity with printed words, digits, and the like to make such materials proper materials for mental testing in these grades.

The tests measure (1) common observation, (2) aesthetic differences, (3) associated objects, (4) discrimination of size, (5) picture parts, (6) picture completion, and (7) dot drawing. From 10 to 15 pupils can be examined at a time, although a skillful examiner may simultaneously test more. Complete directions for giving and scoring the test are given in a *Manual of Directions* accompanying the test.

The test authors report reliability coefficients of 0.88 and 0.93 for the instrument. They report also validity coefficients varying from 0.55 to 0.83 (median coefficient value, 0.78), against such criteria as scores yielded by other individual and group mental tests and teachers' judgments as well as composite ratings of mental ability which involve a number of factors. It should be obvious that this is a good mental test for kindergarten and primary grades.¹

*The McCall Multi-mental Scale.*²—This is a verbal test of 100 items, each including five words. The task is to indicate the one word in each group of five that does not belong with the other four. The content of the test is well illustrated by the following excerpt from the instrument:

1	2	3	4	5
fly	cup	horse	lesson	grass
burn	fork	calf	problem	coal
gasoline	saucer	colt	teacher	carbon
coal	bowl	hen	learn	tar
wood	knife	cow	solve	soot

The type of response demanded is quite simple, and it is uniform throughout the test, making for rapid responding. Despite the apparent simplicity, however, the test items are subtle in their demands. Except in a very general way, no two items employ the same point of view. The pupil is not told the point of view that he must take in order to succeed with a test item; he must divine this for himself. The test author maintains that this

¹ The data on the reliability and the validity of this test have been taken from R. Pintner and B. V. Cunningham, *Manual of Directions and Key for the Pintner-Cunningham Primary Mental Test*, p. 3.

² Devised by W. A. McCall and published by the Bureau of Publications, Teachers College, Columbia University, New York.

divination feature accounts for this test's being the only one that may be used from grade 3 through the university.¹

Five answers are possible to each of the test items, each with some degree of correctness. The individual responding receives a score from 0 to 10 on each item, depending upon the correctness of the word that he marks. The test author says that this refined scoring tends to increase both the reliability and the validity of the test, since each item thus has a chance to divide a group into five ability groups, as is not done by the usual method that marks a test item either right or wrong.² Even though the scorer avails himself of the window stencil furnished with the instrument, scoring is laborious with this test unless a calculating machine is used.

The author provides data as to the validity of the test, using as a criterion composites involving weighted scores yielded by mental tests. The coefficients range from 0.81 to 0.95, with a median value of 0.92. No specific data are given as to the reliability of the instrument.² Grade placement score equivalents to raw scores are available.

It should be noted that the verbal character of this instrument makes it invalid for use with groups of pupils that include individuals with an English-language handicap.

*The Kuhlmann-Anderson Intelligence Tests.*³—This test is the result of 10 years of research, and the end product is really a scale

School grade	Tests	Age where tests best fit
1 (First semester).....	1 to 10	6-0
1 (Second semester).....	4 to 13	6-6
2.....	8 to 17	7-6
3.....	12 to 21	8-6
4.....	15 to 24	9-6
5.....	18 to 27	10-6
6.....	22 to 31	11-6
7 to 8.....	25 to 34	13-0
9 to 12.....	30 to 39	15-6

¹ W. A. McCall, *Directions for Using Forms 1 and 2 of the Multi-mental Scale*, p. 1.

² *Ibid.*

³ Devised by F. Kuhlmann and Rose G. Anderson and published by the Educational Test Bureau, Minneapolis, Minn.

which has been broken up into nine separate tests, each with 10 separate subtests which occur each in two or three different examinations. The separate tests are grouped into batteries for group testing as shown in the table on page 286.

Tests 1 to 14 inclusive and Tests 16, 17, 18 and 20 should be classified as performance tests, although numbers are used in several instances. The numbers are used in concrete situations. In the remaining 17 tests, persons responding are required to deal with words and numbers.

The test authors provide mental-growth units and "per cent of average development" measures as alternates for the more common mental age and intelligence quotient.¹ They have not employed the customary methods of determining and reporting the validity and reliability of the tests. The procedures followed, however, plus data reported elsewhere² indicate that the tests possess good reliability and validity.

Complete instructions for giving the tests are printed in the *Instruction Manual*. On one page, the detailed instructions are given; and on the opposite page, the test proper is reproduced. This makes giving the test easy, since the materials about which the tester is reading are before him at all times. Each test is separately timed.

Scoring keys on cards are furnished for each test separately. On the back of each key are given the mental-age scores and mental-growth units earned by successfully completing each single item in each test. Obviously, each separate test, with 10 subdivisions, yields 10 mental ages, so the mental age for the

¹ F. KUHLMANN and ROSE G. ANDERSON, *Instruction Manual for the Kuhlmann-Anderson Intelligence Tests*, pp. 17-21.

² For example, see F. KUHLMANN, The Pearson Formula, and a Further Note on the Kuhlmann-Anderson Intelligence Tests, *Journal of Applied Psychology*, 13: 32-45 (February, 1929).

KEY. TEST 27

Example:
Do not count

	16	6	26
8	18	10	32
	50	27	35
44	10	9	21
28	25	27	6
	50	21	44
	18	8	35
32	27	28	35
	9	25	16

entire test is taken as the median mental age in the group of 10 mental ages.

Duplicate forms are not available, since the test authors recommend the use of tests immediately above and below those previously given if a more extensive examination of a given pupil is desired.

This instrument appears to be one of the more carefully standardized mental tests available for use in the elementary schools.

*The Otis Quick-scoring Mental Ability Tests.*¹—This series of measuring instruments includes three separate tests known as the Otis Quick-scoring Mental Ability Tests: Alpha, Beta, and Gamma. The Alpha test is designed for use in grades 1 to 4; the Beta test, in grades 4 to 9; and the Gamma Test, in high schools and colleges.

The Alpha test is actually two separate tests, both using the same set of pictures, one verbal and the other nonverbal. One set of directions calls for the marking in each row of a picture that is different from the others in the row, and the other calls for marking in each row a different picture in accordance with specific directions. Either or both the verbal and the nonverbal tests may be given in obtaining the final score. This test is difficult for beginning first-grade pupils, and its use is not recommended until the beginning of the second half of the first year of grade-school work.

The Beta test is a revision and extension of the Intermediate Examination of the Otis Self-administering Tests of Mental Ability. It is a verbal test, with a wide variety of types of items. The feature of this test is the method of response, placing a cross in the circle opposite the test item with the same number as that of the correct answer. When the test is opened up, the circles in which the correct answers should appear lie directly below similar cut-out circles in the scoring key, making it possible to score the test very rapidly with only one application of the scoring key.

The reliability coefficients indicate that the test is sufficiently consistent. For example, reliability coefficients for the Beta test, grades 4 to 9 separately, vary from 0.665 to 0.979; and for the grades combined, the reliability coefficient is 0.96. The average

¹ Devised by A. S. Otis and published by the World Book Company, Yonkers-on-Hudson, N. Y.

of 12 separate grade reliability coefficients is 0.79. The probable error of a score is 2.7 months, indicating that a pupil's score will be in error between 0 and 2.7 points in 50 per cent of the cases.¹ Obviously, the reliability of the test is satisfactory. The author gives no specific data as to the validity of the instrument. Grade-placement scores are available in the *Manual of Directions*.

*The Detroit Intelligence Tests.*²—Actually, there are two Detroit tests, the Detroit Primary Intelligence Test, devised for grades 2 to 4 inclusive, and the Detroit Alpha Intelligence Test, for grades 5 to 9. The primary test includes the following subtests: (1) visual; (2) missing parts, absurdities; (3) auditory information; (4) selection of the largest of several items; (5) word information; (6) series completion; and (7) letter matching. The Alpha test subtests measure (1) general information; (2) vocabulary, opposites; (3) classification; (4) visual imagery; (5) logical selection; (6) analogies; (7) arithmetic, number series; and (8) reading of mixed sentences.

The reliability of the primary test is indicated by an r of 0.93 (229 cases). A validity coefficient of 0.81 against the criterion, Stanford-Binet mental age, for 200 cases is reported.³ Reliability coefficients varying between 0.854 and 0.907 have been reported by the author for the Alpha test. A validity coefficient for 78 cases against Stanford-Binet mental age was found to be 0.653. A validity coefficient of 0.704 was obtained on 83 cases between Detroit Alpha scores and Terman group test scores.⁴ These two tests appear to be sufficiently reliable and valid.

*The California Tests of Mental Maturity.*⁵—Three batteries of these tests are available for use in elementary-school grades: a preprimary battery for use in the kindergarten and in grade 1; a primary battery for use in grades 1 to 3; and an elementary battery for use in grades 4 through 8. Each of these tests is diagnos-

¹ A. S. OTIS, *Manual of Directions for the Beta Test, Otis Quick-scoring Tests*, pp. 10, 11.

² Devised by H. J. Baker and published by the Public School Publishing Company, Bloomington, Ill.

³ H. J. BAKER, *Manual of Directions for the Detroit Primary Intelligence Test*, p. 15.

⁴ H. J. BAKER, *Manual of Directions for the Detroit Alpha Intelligence Test*, p. 16.

⁵ Devised by Elizabeth T. Sullivan, W. W. Clark, and E. W. Tiegs and published by the California Test Bureau, Los Angeles.

tic in type. The analytical comparison of the various sections indicates a definite central factor, yet the same analysis reveals a sufficient specificity for each test to justify its inclusion as a measure of a more or less unique factor.¹ These tests yield each three intelligence quotients, a nonverbal I.Q., a verbal I.Q., and the I.Q. ordinarily obtained by administering a test of mental ability.

The test authors have derived a Short-form of the California Test of Mental Maturity for the use of testers to whom economy of time in administering the tests is of more value than an extensive diagnosis. The discussion below makes clear the differences between the two tests. Space limitations prevent giving in detail all of the information concerning each of the three batteries, both long and short forms. In general, the same general facts have been found for each of the separate series of tests. The discussion hereafter will deal only with the elementary battery.

The first matter to be thought of deals with the content of the test.²

THE DIAGNOSTIC, OR LONG, FORM

These tests are unique in that each series is preceded by tests of visual acuity, auditory acuity, and motor coordination, the purpose of which is to detect those pupils with defects sufficiently serious to prevent obtaining a valid diagnosis of mental maturity with the remainder of the test.

In general, the mental maturity test proper samples memory (immediate and delayed); maturity of apperceptive processes; spacial relationships; and logical and mathematical aspects of reasoning. Certain tests are presented in both verbal and non-language form in order to obtain a separate evaluation of each.

Each series is thus built upon a plan which (1) takes into account recognized mental processes or functions, and (2) corresponds with present knowledge of the nature of mental growth and development.

THE SHORT FORM

This series of tests is unique in that each series is preceded by a pretest of visual acuity, the purpose of which is to detect those pupils with serious defects.

¹ It should be pointed out that this situation applies also to other group tests, e.g., to the Detroit Intelligence Tests.

² This information is taken from the *Manuals of Directions* of the two tests and from the front-cover pages of the tests proper.

In general, the six sections of the mental maturity test proper sample maturity of apperceptive processes; spacial relationships; and logical and mathematical aspects of reasoning. Some of the tests are presented in verbal and others in non-language form in order to obtain a separate evaluation of mental processes with and without use of language symbols.

Each series is thus built upon a plan which (1) takes into account recognized mental processes or functions, and (2) corresponds with present knowledge of the nature of mental growth and development.

A clearer understanding of the contents may be obtained by reviewing briefly the content of the tests. It should be noted that the short-form test includes exactly one-half of the tests to be found in the diagnostic, long form. In the list below naming the

THE DIAGNOSTIC, OR LONG FORM

TEST	FACTOR	Possible Score
A. Memory		44
4. Immediate Recall*		24
5. Delayed Recall		20
B. Spacial Relationships		45
6. Sensing Right and Left*		20
7. Manipulation of Areas*		15
8. Foresight in Spacial Situations*		10
C. Reasoning		95
9. Opposites*		15
10. Similarities*		15
11. Analogies*		15
12. Number Series*		10
13. Numerical Quantity*		10
14. Numerical Quantity		15
15. Argument		15
D. 16. Vocabulary		50
Total Mental Factors		234

THE SHORT FORM

TEST	FACTOR	Possible Score
Non-Language Tests		50
1. Sensing Right and Left		20
2. Manipulation of Areas		15
3. Similarities		15
Language Tests		80
4. Numerical Quantity		15
5. Inference		15
6. Vocabulary		50
Total Mental Factors		130

subtests in the long form, those marked with a star are non-language tests. The preliminary tests are not named in either list.

The test authors report as follows concerning the reliability of these tests.¹

THE DIAGNOSTIC, OR LONG FORM

Needless to say, the number and variety of test situations assure a high reliability. The test data for 1,048 pupils in representative school districts have been analyzed and reliability coefficients have been computed by the split-halves method and corrected by the Spearman-Brown formula. The reliability of the Total Mental Factors score and the major test items are as follows:

	Grades					
	4	5	6	7	8	4-8
Total mental factors.....	0.956	0.950	0.958	0.962	0.963	0.967
Language factors.....	0.930	0.946	0.944	0.944	0.956	0.958
Non-language factors.....	0.925	0.922	0.932	0.949	0.919	0.946
Test A. Memory.....	0.906	0.933	0.926	0.920	0.943	0.936
Test B. Spacial relationships.....	0.873	0.898	0.900	0.894	0.878	0.904
Test C. Reasoning.....	0.879	0.871	0.912	0.928	0.894	0.924
Test D. Vocabulary.....	0.912	0.922	0.901	0.921	0.942	0.936
No. of pupils.....	144	270	308	192	134	1048
S.D. (M. A. in months).....	12.0	13.0	13.5	14.0	15.5	15.0

The probable error of estimate for an individual pupil varies from two months mental age when the reliabilities for single grade are .95 and above to four months when the reliabilities are .87 to .89.

THE SHORT FORM

The reliability of the short-form test has been computed ($N = 1,048$ pupils) by the split-halves method and corrected by the Spearman-Brown formula, with the results as follows:

¹ ELIZABETH T. SULLIVAN, W. W. CLARK, and E. W. TIEGS, *Manual of Directions for the California Test of Mental Maturity—Elementary Battery*, p. 1.

Manual of Directions for the California Short-form Test of Mental Maturity, Elementary S-form, p. 1.

	Mental ages		Intelligence quotients	
	<i>r</i>	P.E. est.	<i>r</i>	P.E. est.
Total mental factors	0.946	4.8 mo.	0.941	3.9 I.Q. points
Language tests.....	0.942	5.0 mo.	0.910	4.5 I.Q. points
Non-language tests.....	0.880	6.9 mo.	0.853	5.6 I.Q. points
Standard deviation.....	20 months		16 I.Q. points	

In determining the reliability of measurement for an individual pupil, the probable error of estimate (P.E. est.) may be used. As shown by the foregoing data, the probable error of estimate for total mental factors M.A. is 4.8 months, and I.Q. is about 4 points.

The test authors say concerning the validity of the test:

The validity of any mental test is difficult to establish; there are no purely objective criteria or standards which correspond to the factors or abilities in terms of which conceptions of mentality are currently described.

The authors of these tests believe that the multiple-factor theory of intelligence comes closer to explaining observable phenomena than does the strong central-factor theory alone. They recognize the importance of philosophical contributions, but they believe that progress in determining the nature of mentality and the value of tests of mental maturity is dependent largely upon further studies in factor analysis which employ analytical and statistical techniques. This series of tests recognizes contributions already made by including samplings of verbal ability, mathematical ability, spacial relations, and logic. The existence of all these as virtually independent and distinct factors is gaining acceptance. The traditional method of correlating the results of this battery with the averages of other intelligence tests (protecting results by observing the usual cautions regarding sampling and other statistical safeguards) reveals that the Total Mental Factors I.Q.s obtained with this test may be used for comparative purposes. However, dealing only with mental ages and intelligence quotients obscures and ignores the separate important factors which constitute mentality; and it is in terms of these factors that the abilities of children should be diagnosed. This battery attempts such a diagnosis on the basis of present knowledge; further evidence of validity must await further knowledge of the nature of mental maturity.¹

¹ ELIZABETH T. SULLIVAN, W. W. CLARK, and E. W. TIEGS, *Manual of Directions for the California Test of Mental Maturity—Elementary Battery*, pp. 1, 2.

The authors have worked out intercorrelations and factor-analysis data dealing with the test, evidencing that the several subtests actually do measure different aspects of mental ability.

Before continuing with the discussion, it is desirable to point out the significance of the preceding discussion, in so far as it has dealt with the content of group tests of mental ability. No effort has been made to advance the claims of any specific test; rather, an effort has been made to protect the content and procedures of all tests, using only as much as is absolutely necessary to give the reader a rather general knowledge of mental tests that have been devised for use in elementary grades. The person who would possess detailed, specific information concerning any single test or group of tests of mental ability must obtain the instrument or instruments, together with the following materials: the manual for the test, the scoring keys, and all added data that the authors can furnish. In addition to this, an examination should be made of the literature in professional publications to obtain such data as may have been printed concerning the instrument or instruments.

The Validity of Tests.—Although the method of intercorrelating total scores yielded by several mental tests in an effort to establish the validity of the instruments has been somewhat discredited, as has been indicated in the preceding discussion, it still is used. It would seem undesirable to leave the matter without presenting some evidence dealing with it. Data are given from two studies, with mental ages used in both instances.

The information given in Table XXIII involves scores yielded by the following tests: the Stanford-Binet; the Otis Group Intelligence Scale, Primary Examination, Form A; the Rhode Island Intelligence Test, Form A; the Kingsbury Primary Group Intelligence Scale, Form A; the Pressey Primary Classification Scale, Form A; the Haggerty Intelligence Examination, Delta I; and the Pintner-Cunningham Primary Mental Test. One hundred pupils in grade 1 responded to these tests.¹

It should be noted that these coefficients are reasonably consistent in value, if the Rhode Island test is eliminated from consideration. Furthermore, the only test for which no single coefficient falls below 0.60, the desirable minimum limit for

¹ The data were taken from P. V. SANGREN, *Validity of Primary Intelligence Tests*, *Journal of Applied Psychology*, 13: 394-412 (August, 1929).

validity, is the Pintner-Cunningham test. The Stanford-Binet is an individual test. Six of the group tests show all of their intercorrelations above 0.60, with the actual values lying between 0.603 and 0.789. These facts indicate that these group tests measure whatever they do measure uniformly to a very marked degree.

TABLE XXIII.—INTERCORRELATIONS BETWEEN SCORES YIELDED BY EIGHT MENTAL TESTS
N = 100, Grade 1 Pupils

	Pressey	Otis	Haggerty	Kingsbury	Rhode Island	Detroit	Pintner-Cunningham
Stanford-Binet.....	0.672	0.631	0.575	0.579	0.432	0.521	0.655
Pressey.....		0.789	0.729	0.658	0.469	0.688	0.643
Otis.....			0.761	0.773	0.513	0.733	0.759
Haggerty.....				0.643	0.545	0.697	0.753
Kingsbury.....					0.433	0.603	0.642
Rhode Island.....						0.627	0.639
Detroit.....							0.732

The second analysis made use of four mental tests: the Henmon-Nelson Tests of Mental Ability, Test for Grades 3 to 8, Form A; the McCall Multi-mental Test, Form I; the California Mental Maturity Test, Elementary Battery; and the Otis Quick-

TABLE XXIV.—INTERCORRELATIONS BETWEEN SCORES YIELDED BY FOUR MENTAL TESTS
N = 100, Grade 7

	Otis II	Henmon-Nelson I	Henmon-Nelson II	McCall I	McCall II	California I	California II
Otis I.....	0.70	0.58	0.57	0.52	0.43	0.71	0.53
Otis II.....		0.66	0.74	0.66	0.55	0.66	0.69
Henmon-Nelson I.....			0.94	0.60	0.63	0.60	0.80
Henmon-Nelson II.....				0.64	0.70	0.68	0.69
McCall I.....					0.74	0.50	0.58
McCall II.....						0.52	0.51
California I.....							0.93

NOTE.—The Roman numerals indicate the first and second administration of one form of each test. This table includes both reliability and validity coefficients.

scoring Mental Ability Tests, Beta Test, Form A. One hundred pupils in grade 7 served as subjects for this study. The inter-correlations obtained by using the mental-age scores yielded by the tests are presented in Table XXIV.¹

Two of the four tests, the Henmon-Nelson and the California, have high reliability coefficients and a majority of their validity coefficients above 0.60. Seven of the 12 McCall test validity coefficients are below 0.60; and 6 of the 12 Otis test validity coefficients, below 0.60. The reliability coefficients of these two tests also are low. There is not the uniformity in measurement among these four tests that exists among six of the eight tests considered in Table XXIII.

Scores Yielded by Mental Tests.—The score yielded by a group mental test ordinarily is a raw, or point, score. The test makers ordinarily provide mental-age equivalents and grade-placement score equivalents to the raw scores. These derived scores are based on the performance of pupils examined during the standardization of the test. The intelligence quotient is computed using the life age and mental age of a given pupil.

A portion of the norms table for one mental test appears on page 297.²

The manner in which this table is used may be made clear by means of an illustration. A boy exactly fifteen years of age responds to the California Test of Mental Maturity—Elementary Battery—and earns a score of 146 points. Referring to the norms table, this score is found to indicate a mental age of 12-0. Since the life age of the pupil is 15-0, the I. Q. is 80. Mental age and the intelligence quotient remain the most commonly used derived scores yielded by mental tests. Various test makers have utilized other measures, such as units of mental growth, grade-placement scores, and indices of brightness, but as yet these have not been used widely.

It should be pointed out that the intelligence quotient is a qualitative measure, whereas mental age is quantitative. For

¹The data were taken from C. McCARTHY, *A Statistical Analysis of Standardized Mental Tests for Upper Grade Elementary Pupils*, p. 38, unpublished Master's thesis, Sul Ross (Texas) State Teachers College, 1939.

²This is a part of the norms table for the California Test of Mental Maturity—Elementary Battery. The complete table is given on p. 12 of the *Manual of Directions* for the test.

this reason, the latter should always be used in evaluating pupils' ability to learn. The only function of the I. Q. is to predict

Mental age		<i>E</i> Total mental- factors score	<i>F</i> Language score	<i>G</i> Non- language score	Grade placement	Age, in months
Months	Yr.-mo.					
144	12-0	145-146	47	98	6.4	144
145	12-1	147-148	48	99	6.5	145
146	12-2	149-150	49	100	6.6	146
147	12-3	151-152	50-51	...	6.6	147
148	12-4	153-154	52	101	6.7	148
149	12-5	155-156	53	102	6.8	149
150	12-6	157-158	54	103	6.9	150
151	12-7	159-160	55	104	6.9	151
152	12-8	161-162	56	105	7.0	152
153	12-9	163-164	57	106	7.1	153
154	12-10	165-166	58	107	7.2	154
155	12-11	167-168	59	108	7.3	155
156	13-0	169-170	60	109	7.4	156
157	13-1	171-172	61	...	7.5	157
158	13-2	173-174	62	110	7.6	158
159	13-3	175-176	63-64	...	7.7	159
160	13-4	177-178	65	111	7.8	160
161	13-5	179	66	...	7.9	161
162	13-6	180	67	112	8.0	162
163	13-7	181	68	...	8.1	163
164	13-8	182	69	...	8.2	164
165	13-9	183	70	113	8.3	165
166	13-10	184-185	71	...	8.4	166
167	13-11	186	72	...	8.5	167
168	14-0	187	73	114	8.6	168
169	14-1	188	74	...	8.7	169
170	14-2	189-190	115	8.8	170
171	14-3	191	75	...	8.8	171
172	14-4	192	76	116	8.9	172
173	14-5	193	117	9.0	173
174	14-6	194	77	...	9.1	174
175	14-7	195	118	9.2	175
176	14-8	196-197	78	...	9.3	176
177	14-9	198	79	...	9.4	177
178	14-10	199	119	9.5	178
179	14-11	200	80	...	9.6	179
180	15-0	201	81	120	9.7	180

mental age at a future time. If this measure is to serve effectively for obtaining mental age, it must be relatively constant.

The next portion of this chapter will be concerned with the constancy of the I. Q.

Constancy of the I. Q.—Numerous studies have been made of the constancy of the I. Q., and all have pointed to the fact that it remains relatively constant from the time of one test to that of the next, whether the time interval between the two tests is two weeks or five years.¹ Typical data are assembled in Table XXV. The information agrees with that given by other writers.

TABLE XXV.—COEFFICIENTS OF CORRELATION INDICATING THE CONSTANCY OF THE I. Q.

Author	Year	N	r	Test used	Interval between tests
Broom.....	1927	50	0.86	Terman	6 to 23 months
Broom.....	1927	102	0.93	Terman	1 to 23 months
Broom.....	1930	212	0.83	Otis Self-Administering	5 months
Lauderbach and Porter.....	1931	193	0.72	Pintner-Cunningham	6 to 17 months
Lauderbach and Hause.....	1932	150	0.79	McCall	12 months
Randall.....	1927	152	0.79	Stanford-Binet	0 to 66 months

It will be recalled that certain test authors indicate that the probable error of a score on certain of the group tests, *i.e.*, the amount of error existing between true and obtained scores in 50 per cent of the cases, varies from two to four months mental age. This is about the amount of change in mental-age placement due to median changes, both positive and negative, in intelligence quotients between tests and retests. From this statement, and from the evidence in Table XXV, it is obvious that the intelligence quotient is not absolutely constant, but it should be equally obvious that it is so nearly constant that it can be used safely for prediction of mental age for time intervals of one to three years, depending on the test that has been used. It should be noted that primary mental tests yield less constant I. Q.s than do tests for intermediate and upper grades.

¹ For additional information on the constancy of the I. Q., see V. E. DICKSON, *Mental Tests and the Classroom Teacher*, p. 66, Yonkers-on-Hudson, N. Y.: World Book Company, 1923.

F. N. FREEMAN, *Mental Tests*, p. 345.

R. PINTNER, *Intelligence Testing—Methods and Results*, pp. 89, 90.

Uses of Mental Tests.—Scores yielded by mental tests are the best prognostic measures of school success yet discovered. The function of these measures is to aid the teacher to adjust the work of the classroom to the ability of her pupils. This may be accomplished by classifying pupils properly on their entrance into a class group. If, for example, a pupil with a high mental age is entered in a class in which the work presents no challenge to him, other factors being equal, he should be advanced to a higher grade level. Grouping within a class, so that pupils with similar abilities work together, is another means of adjusting the work of the school to the pupils' abilities. This is particularly effective in primary grades. Still another use is in determining whether or not pupils are able to do certain types of work, *e.g.*, determining reading readiness of pupils in grade 1.¹ Finally, mental-test data are of value in attempting to solve childrens' problems and the difficulties of problem pupils. If the adjustment of the school to the needs and abilities of each child is to be made, the teacher must study each child as an individual and from every possible viewpoint. She must place the child so that he will have an opportunity to learn, and then she must teach, diagnose the pupils' learning, and reteach in the light of the diagnosis. Furthermore, she must guide the developing personality of the child in desirable directions. With all these things to accomplish, it becomes evident that test findings, and particularly mental-test results, are highly important to the teacher attempting conscientiously to do good professional work.

Summary.—This chapter has attempted to present information dealing with the nature, content, and uses of tests of mental ability and of the scores that these instruments yield. The discussion has been limited to group tests only because they require little specialized training in their administration and use. Obviously, it has been possible to present only the minimum information in so brief a discussion, and the serious student is urged to read widely and to study actual test materials extensively and thoroughly in order to obtain an adequate knowledge of the facts of mental tests and testing.

Readings for Further Study

NOTE.—Periodical articles have been omitted, primarily because the quantity of such material is so large that it is impossible to present a full,

¹ M. LUCILLE HARRISON, *Reading Readiness*, p. 6.

300 MEASUREMENTS IN THE ELEMENTARY SCHOOL

representative list in brief compass. An effort has been made in the Questions and Exercises to guide the student to periodical materials.

1. B. R. BUCKINGHAM, *Research for Teachers*. Chap. III.
2. W. F. DEARBORN, *Intelligence Tests*. Chaps. III, IV.
3. V. E. DICKSON, *Mental Tests and the Classroom Teacher*. Chaps. III, VI-X.
4. N. FENTON and D. A. WORCESTER, *Educational Measurements*, Chap. II.
5. F. N. FREEMAN, *Mental Tests*. Chaps. II, VI, VII.
6. C. L. HULL, *Ability Testing*. Chap. III.
7. F. KUHLMANN, *A Handbook of Mental Tests*. Chaps. I-III. Baltimore: Warwick and York, Inc., 1922.
8. R. PINTNER, *Intelligence Testing—Methods and Results*. Chaps. VII-X, XIV-XVI, XXI.
9. S. L. PRESSEY and L. C. PRESSEY, *Introduction to the Use of Standardized Tests* (rev. ed.). Chaps. XI, XII.
10. H. L. SMITH and W. W. WRIGHT, *Tests and Measurements*. Chap. XIX.
11. M. R. TRABUE, *Measuring Results in Education*. Chaps. XIV, XVIII, XIX.
12. L. W. WEBB and A. M. SHOTWELL, *Standard Tests in the Elementary School*. Chaps. V-VII.
13. F. L. WELLS, *Mental Tests in Clinical Practice*. Chaps. III, IV, VI. Yonkers-on-Hudson, N. Y.: World Book Company, 1927.
14. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chaps. XV, XVI.

The following two bibliographies list references from which data in Tables XXII and XXV were obtained.

Bibliography for Table XXII

1. B. H. BROOM, A Note on the Relationship between Mental Ability and School Achievement, *El Paso Schools Standard*, 14 (No. 1): 14-17 (October, 1936).
2. M. E. BROOM, Measuring Mental Ability in the Intermediate Grades of the Elementary School, *School and Society*, 35: 323-324 (1932).
3. F. W. DESILVA, A Study of the Measures Used in Predicting Academic Achievement in Grade 7 in a Junior High School, p. 108, unpublished Master's thesis, University of Southern California, 1933.
4. A. M. JORDAN, in T. L. KELLEY and A. C. KREY, *Tests and Measurements in the Social Studies*. P. 458.
5. R. C. PERRY, *A Group Factor Analysis of the Adjustment Questionnaire*. Pp. 27, 29. Los Angeles: University of Southern California Press, 1934.
6. P. A. WITTY and J. F. TAYLOR, Some Results of the Multi-mental Test, *Journal of Educational Psychology*, 20: 299-303 (1929).
7. L. A. WRIGHT, A Study of Certain Personality Traits that Reinforce or Interfere with the Function of Mental Ability to Affect School Achievement, p. 28, unpublished Master's thesis, University of Southern California, 1931.

Bibliography for Table XXV

1. M. E. BROOM, Constancy of the I.Q., *School and Society*, 25: 295-296 (Mar. 5, 1927).
2. M. E. BROOM, A Study of the Constancy of the I.Q., *Educational Research Bulletin (Los Angeles)*, 7 (Nos. 2, 3): 2-3 (October-November, 1927).
3. M. E. BROOM, How Constant Is the I.Q. Yielded by the Otis Self-administering Test of Mental Ability? *Journal of Educational Research*, 22: 53-55 (June, 1930).
4. J. C. LAUDERBACH and M. POWELL PORTER, On the Constancy of the I.Q., *School and Society*, 33: 675-676 (May 16, 1931).
5. J. C. LAUDERBACH and ENID HAUSE, On the Reliability and Validity of Derived Scores Yielded by the McCall Multi-mental Scale, *Journal of Applied Psychology*, 16: 322-323 (June, 1932).
6. F. B. RANDALL, A Study on the Constancy of the I.Q., *School and Society*, 26: 311-312 (Sept. 3, 1927).

Questions and Exercises

1. Prepare a series of six reports of periodical articles dealing with mental tests and testing. Follow the directions given in Exercise 9, Chapter VI (page 140).

2. Four tests of mental ability were given to 69 pupils in grade 8 in a public school. The intercorrelations based on mental-age scores yielded by the tests are given in the table below. Interpret these data.

	Otis S-A	Detroit	Terman Group Test of Mental Ability
Kuhlmann-Anderson Intelligence Test...	0.68	0.62	0.68
Otis Self-administering Test of Mental Ability.....	0.76	0.74
Detroit Alpha Intelligence Test.....	0.71

3. In the information below, two of three measures are given (life age, mental age, and intelligence quotient). Compute the missing item, and determine grade placement, using the partial table of norms for the California Test of Mental Maturity—Elementary Battery—given in this chapter (page 297).

CA	MA	IQ	G.P.
12-4	_____	102	_____
_____	13-7	111	_____
13-2	12-8	_____	_____
14-3	13-11	_____	_____
_____	14-8	96	_____
12-5	_____	119	_____

4. What are the chief difficulties encountered when attempting to measure the mental ability of children entering primary grades?

5. A child has a mental age of 10-2 when his life age is 8-1. Assuming conditions of consistent mental maturation, what will his mental age be when his life age is 12-3?

6. How can a group mental test be made useful as an aid in educational guidance?

7. Why are group mental tests used in schools in preference to individual tests of mental ability?

8. Who in a school system should be held responsible for the selection and administration of mental tests and for the subsequent uses of pupils' responses?

9. What objections are there against grouping pupils on the basis of mental age only?

10. Is the intelligence quotient of any value for grouping purposes? Why?

11. Write a brief and careful presentation of the limitations existing in present group mental tests.

12. Establish your own set of criteria for a good group mental test. Justify these criteria. Mention several available group mental tests that possess one or more of these criteria.

13. What process does the standardization of a group mental test involve?

14. Do intelligence quotients yielded by different group mental tests for a given pupil always have the same value? Why?

15. In what ways may the validity of a group mental test be established?

16. What types of exercises usually are included in primary mental tests? In mental tests devised for use in intermediate and upper elementary grades?

17. What is the difference between the validity of a group mental test and its reliability? Which of these presupposes the other?

18. What is the present view of psychology as to the constancy of the intelligence quotient?

19. To what extent do practice effects affect mental-test scores on successive administrations of duplicate forms of the same test? What effect does practice in taking mental tests have on a pupil's efficiency in taking mental tests that he has not attempted previously?

20. Are nonverbal group mental tests used extensively in intermediate and upper elementary grades?

21. Is a year of mental retardation equally serious at all life ages? Why?

22. How do you explain the high correlation existing between derived scores yielded by mental tests and batteries of school achievement tests? Why do you think this correlation is higher than that between mental-test scores and teachers' marks?

23. Do intelligence tests measure mental ability directly? What fundamental assumptions underlie the measurement of mental ability?

24. Prepare a tabulation listing all of the specific ways in which mental-test results would be of value in an elementary school.

EXAMINATION ON MENTAL TESTS AND TESTING

PART I

Directions: Below are a number of statements. Some are true; some are false; and some are neither entirely true nor entirely false. Place a check on the line before the letter T at the left of each true statement; before the letter F at the left of each false statement; and before the letter N at the left of each statement that is neither entirely true nor entirely false. If you do not know which response is correct, skip the item, and continue at once with the next one. DO NOT GUESS! Work as rapidly and as accurately as you can.

- | | | | |
|----------------------------|----------------------------|----------------------------|---|
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 1. When a pupil's mental age is 15-4 and his life age 14-8, his I.Q. is 107. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 2. The I.Q. level 110-119 is assigned to normal mentality. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 3. Intelligence tests are useful in educational guidance, but they alone will not suffice for all guidance purposes. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 4. Intelligence tests have done much to show that all children are not created free and equal with respect to their mental abilities. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 5. The upper limit of feeble-mindedness should be drawn at I.Q. 60. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 6. Many recent studies have demonstrated that the I.Q. is constant. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 7. Mental ability, in a large population, is distributed normally, <i>i.e.</i> , in accordance with the law of normal probability. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 8. There is positive and high correlation between scores yielded by mental tests and measures of academic success. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 9. Mental tests are used in diagnosing academic failures in schools. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 10. If a given pupil is tested with a number of different mental tests, the examiner will obtain the same I.Q. on all of them. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 11. The measure of the validity of a mental test can be obtained only in indirect ways. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 12. Mental ability can be measured directly. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 13. Authorities differ considerably in their concepts of intelligence. |
| <input type="checkbox"/> T | <input type="checkbox"/> F | <input type="checkbox"/> N | 14. The California Test of Mental Ability includes two types of tests, a diagnostic one and a short-form for survey purposes. |

- _____ T _____ F _____ N 15. Terman writes that intelligence is the ability to learn.
- _____ T _____ F _____ N 16. Hull defines intelligence as progressive adaptability.
- _____ T _____ F _____ N 17. The material contained in mental tests is always particularized, and the instruments therefore are to some degree specialized tests.
- _____ T _____ F _____ N 18. Spearman's *g*-factor alone cannot account for intellectual achievement and for mental-test scores.
- _____ T _____ F _____ N 19. Mental traits are dependent upon anatomical and neurological structures.
- _____ T _____ F _____ N 20. Talented persons not only start with greater initial capacities but seem also to be more capable of intensive application, profiting more both relatively and absolutely than the less gifted.
- _____ T _____ F _____ N 21. Current tests of mental ability call into exercise the power to deal with new situations.
- _____ T _____ F _____ N 22. Present mental tests measure only two characteristics of mental ability: speed of work and the difficulty level of the tasks.
- _____ T _____ F _____ N 23. All persons reach mental maturity at the same life age.
- _____ T _____ F _____ N 24. The range of values of validity coefficients for modern mental tests is between positive coefficients of correlation of 0.50 and 0.70.
- _____ T _____ F _____ N 25. When a given pair of mental abilities are to any extent correlated with each other, to this degree they may be considered as depending mutually upon a common factor.
- _____ T _____ F _____ N 26. The data as to the life age at which mental maturity is achieved are scanty.
- _____ T _____ F _____ N 27. All of the tasks of the Otis Quick-scoring Mental Ability Tests are verbal.
- _____ T _____ F _____ N 28. The Pintner-Cunningham Primary Mental Test is a nonverbal test for kindergarten and first grade pupils only.
- _____ T _____ F _____ N 29. Spearman suggests that *g* measures something in the nature of an energy.
- _____ T _____ F _____ N 30. The unit-factor theory of intelligence recognizes intelligence as a single variable.

- _____ T _____ F _____ N 31. The existing psychological theories as to the nature of intelligence are clearly distinct in their neurological implications.
- _____ T _____ F _____ N 32. Coefficients of correlation for paired measures of social and academic intelligence usually range between positive 0.40 and 0.60.
- _____ T _____ F _____ N 33. In studies of the constancy of the I.Q., the coefficient has been found to exceed positive 0.80.
- _____ T _____ F _____ N 34. Early attempts to measure intelligence did not succeed, because they depended on the measurement of physical traits.
- _____ T _____ F _____ N 35. In giving mental tests, the personality of the examiner is the governing factor in determining the attitude of the subject.
- _____ T _____ F _____ N 36. M.A. indicates rates of mental growth.
- _____ T _____ F _____ N 37. Great care must be exercised in interpreting mental ages and intelligence quotients of children over thirteen years chronologically.
- _____ T _____ F _____ N 38. Mental level and brightness have two quite different meanings in describing the ages when mental growth is most rapid.
- _____ T _____ F _____ N 39. No general tests of intelligence have yet been devised.
- _____ T _____ F _____ N 40. Few studies of the relationship between social and mechanical intelligence have been made.
- _____ T _____ F _____ N 41. What is measured by current mental tests includes many different mental functions, each more or less uncorrelated with the others.
- _____ T _____ F _____ N 42. The intelligence quotient is a derived score.
- _____ T _____ F _____ N 43. Mental age and mental ability have two quite different meanings.
- _____ T _____ F _____ N 44. It is possible by coaching to raise a child's score on a mental test.
- _____ T _____ F _____ N 45. Unless the subject can read, it is impossible to secure a valid measure of a child's mental ability.
- _____ T _____ F _____ N 46. The Cole-Vincent Primary Mental Test is a nonverbal test for kindergarten, first-, and second-grade children.
- _____ T _____ F _____ N 47. The McCall Multi-mental Scale is suited for use in grades 2 to 9.

- _____ T _____ F _____ N 48. The term psychometric refers to the measurement of intelligence in its broadest sense.
- _____ T _____ F _____ N 49. A pupil whose M. A. is 12-6 and whose C. A. is 15-6 has an I. Q. of 82.
- _____ T _____ F _____ N 50. The discovery of a valid and reliable test of mental ability has been largely an empirical, trial-and-error affair.

PART II

Directions: In each exercise below, the unfinished statement is completed by one of the four items immediately following it. You are to place the number of the item that completes the exercise correctly on the line at the left of the exercise number. Work as rapidly and as accurately as you can.

- _____ 1. The Pintner Non-language Primary Mental Test may be used in (1) grades K, 1, and 2. (2) grades K and 1. (3) Kindergarten only. (4) grades 1 and 2.
- _____ 2. The Binet-Simon Scale was first introduced into the United States by (1) Terman. (2) Goddard. (3) Thorndike. (4) Cattell.
- _____ 3. The I. Q. of a pupil whose mental age is 15-2 and whose life age is 12-8 is (1) 118. (2) 120. (3) 112. (4) 125.
- _____ 4. A child whose I.Q. is 118 is classified approximately as having (1) genius or near genius. (2) very superior intelligence. (3) normal intelligence. (4) superior intelligence.
- _____ 5. Validity coefficients based on total scores yielded by tests of mental ability for pupils in single grade groups usually range between positive (1) 0.40 and 0.60. (2) 0.60 and 0.70. (3) 0.50 and 0.75. (4) 0.55 and 0.70.
- _____ 6. The definition "general intelligence is 'a kind of average of all possible aptitudes'" expresses the opinion of (1) Spearman. (2) Koffka. (3) Binet. (4) Hull.
- _____ 7. The person who defined intelligence as "progressive adaptability" is (1) Thorndike. (2) Stern. (3) Porteus. (4) Ballard.
- _____ 8. The theory that mental ability involved in doing an intellectual operation may be resolved into two factors g and s was advanced by (1) Köhler. (2) Thorndike. (3) Thurstone. (4) Spearman.
- _____ 9. Coefficients of correlation defining the relationship between academic or abstract intelligence and mechanical intelligence vary between positive (1) 0.25 and 0.68. (2) 0.10 and 0.54. (3) 0.00 and 0.28. (4) 0.02 and 0.34.
- _____ 10. The inventor of the I. Q. ratio was (1) Stern. (2) Porteus. (3) Lashley. (4) Thorndike.

PART III

Directions: In each exercise below, write the word or words needed to complete the statement correctly. Work as rapidly and as accurately as you can.

1. Manuals of directions accompanying group mental tests give specific instructions as to how the tests are to be _____ and _____; and _____ and _____ norms.

2. Ogden, presenting the viewpoint of _____ psychology, describes intelligence in terms of articulateness of pattern.

3. Four psychological theories as to the nature of intelligence have been advocated. These are (1) _____; (2) _____; (3) _____; and (4) _____.

4. Two of the four psychological theories seem to be more significant than are the others. They are those advocated by _____ and _____.

5. _____ is a quantitative measure of mental ability; _____, a qualitative measure.

6. The following mental tests are useful in elementary grades. Before the name of each test, place the initial K (for kindergarten) or the number for each grade in which the test is useful.

- _____ Terman Group Test of Mental Ability
- _____ The Cole-Vincent Primary Mental Test
- _____ The Detroit Alpha Intelligence Test
- _____ The Kuhlmann-Anderson Intelligence Test
- _____ The Otis Quick-scoring Mental Tests, Beta Test
- _____ The McCall Multi-mental Scale
- _____ The Pintner-Cunningham Primary Mental Test
- _____ The National Intelligence Test

7. Thorndike suggests that quality of intellect depends upon (1) _____ and (2) _____.

8. A number of specific types of intelligence have been identified. They are:

9. _____ characteristics describe the better group tests of mental ability. They are:

10. There are _____ roots from which present mental tests have descended. They are:

CHAPTER XIII

USES OF EDUCATIONAL MEASUREMENTS

It has been pointed out earlier in this volume that many persons view the measurement movement with suspicion and antagonism. To them measurement is tinged with the occult and shrouded in mystery. Such persons cannot comprehend the practical uses of tests, and therefore they regard them as useless and their use as an unnecessary burden on the classroom teacher. The purpose of measurement is to give objective, unbiased, and accurate evidence concerning something about which one desires or needs to know. There is nothing inherently strange about objective and standardized mental and achievement tests. They are based on materials and methods familiar to everyone.

Earlier in this volume, some attention was given to the uses of tests, but the discussion dealt with the purposes rather than specific uses of measurements and measuring instruments.¹ It is the purpose of this chapter to summarize and to amplify these earlier discussions and to give specific examples of practical uses of achievement and mental tests in educational work. The discussion that follows is grouped under three headings: (1) administration, (2) supervision, and (3) classroom teaching.

Administrative Uses.—The administration of a school system is responsible for the establishment, operation, and maintenance of an educational program. Tests, therefore, are essential in the work of school administration, providing information essential in guiding the educational work of the schools.

Data yielded by batteries of school achievement tests show clearly the mass picture of educational achievement in the school system in all grades and subjects. This permits the evaluation of the general efficiency of instruction and the making of comparisons of achievement in the system with that of other systems as well as within the system by schools, grade levels, and class groups. It is possible to prepare summary charts by subjects for

¹ For example, see the later portion of Chap. I, and the introductory paragraph of Chap. VI.

all schools, showing all grade levels but no individual school or class groups, as was done in Fig. 6 (Chapter IV, page 67). When such charts are available for all subjects, the administrator can readily identify the strengths and weaknesses of the school instruction system. This provides a basis for modifying the instructional program, if necessary, and for guiding the work of supervision in improving instructional efficiency. Similarly, such charts can be prepared for each individual school for all subjects, grade levels, and class groups. The information so graphed shows when an instructional strength or weakness is uniform in the system and when it is confined to certain schools. This helps supervision to function where it is needed.

The latter type of summary chart aids the administrator in personnel work. If the charts over a period of years show a certain teacher always below normal in the achievement gains of her pupils during a definite period of instruction, such as one school year, while mental-ability summaries indicate that she is not always teaching retarded pupils, it follows that her teaching is weak in some manner. The supervisory staff must then attempt definitely to improve the work of this teacher. Ultimately, if there is no improvement in pupil achievement, regardless of the assistance that has been given, the data justify her discharge.

There is another aspect of personnel work in which measurements are of material assistance. If a teacher's work shows strength in all subjects but one, *e.g.*, arithmetic reasoning, in terms of her pupils' achievements over a period of time, it is often possible to place her between two teachers who are strong in this particular subject. In this way, over a period of years, the pupils will not suffer through the weakness of a single teacher. Furthermore, the strengths and weaknesses of teachers, judged by the achievements of their pupils, can be taken into account in the making of transfers from school to school, from grade to grade, and from subject to subject where departmentalized instruction is in operation. In the long run, this is advantageous both to the teachers and to the pupils.

The administration can scan achievement results to see whether or not there is any weakness in instructional efficiency affecting any particular group of pupils. For example, Problem 3, Chapter IV (pages 75-76), presents a study of achievement by two racial-backgrounds groups of pupils, one with a language handicap. In

this instance, the problem before the administration was whether or not different teaching methods should be evolved for the handicapped group.

The administration is definitely charged with the responsibility for informing the public concerning its schools. Measurements yield only a part of the information that should be given to the taxpayers of the community, but they do provide facts that should be given in an accounting to the public. Taxpayers and parents are entitled to know whether or not their educational dollar is actually buying a normal dollar's worth of education for the children of the community. In the last analysis, this means that the public should be told whether or not the pupils in the schools are progressing normally in achievement in all subjects and grades.

The administration of a school system is interested also in other things in which measurement is involved, most of which are delegated directly to subordinates, possibly coming to the attention of an administrator only in problem situations. These include (1) the proper classification and promotion of pupils, (2) the selection of textbooks, (3) the proper guidance of pupils, and (4) experimentation with teaching methods, particularly with a view to improving them. These will be treated later in the chapter. It should be obvious that only the major uses of measurement have been presented in the preceding paragraphs and equally obvious that school administrators must study data yielded by achievement and mental tests if they are to function efficiently in furthering educational opportunities for children.

Supervisory Uses.—Administrators plan broadly as to instruction in a school system, and supervisors attempt to carry out these plans in specific ways and in specific situations. The supervisory staff must translate theory into practice, providing the research and service that are essential in carrying out an effective instructional program. Measurement helps to keep supervision impersonal and efficient.

Usually, the administration indicates to the supervisory staff the schools, school grades, and subjects in which attention is needed. Plans should be based on a study of summary test data, which should be made available to the supervisors. In turn, the supervisors should make such further analyses of the test results as might be necessary. Error-analysis charts for class groups

selected at random might be made in an effort to identify common weaknesses of pupils in schools throughout the system.¹ If this procedure shows definite weaknesses in the student population, the supervisors can then plan remedial teaching procedures and devise teaching materials, such as drills, which classroom teachers can generally use in improving achievement in the specific pupil weaknesses. If, however, the results vary as to specific weaknesses from class group to class group, the problem is more difficult. Further tests of a diagnostic character should be given, and further analyses made. An example of such an analysis is given on page 312.²

The types of questions indicated in the chart are: Type I, those which require the pupil to recognize an incident that is essential to the story purpose of the narrative; Type II, those which require the pupil to recognize an incident essential to the development of a part of the narrative; and Type III, those which require the pupil to identify the specific details through which the story is developed. An analysis of the chart shows that the group measured made a larger percentage of errors on Type II than on either Types I or III and a larger percentage of errors on Type III than on Type I. Furthermore, the pupils in the upper half of the class in comprehension made the majority of their errors on Types II and III questions, whereas those in the lower half of the class in comprehension made a better record on Type III questions than on Type I. The more able readers comprehended the story but failed to grasp the supporting parts of the narrative and the details, whereas the less able readers missed the author's purpose and the main points of the story.

Such an analysis chart tells the supervisor and the classroom teacher also (1) the needs of the class as a whole with regard to the three types of reading comprehension; (2) the types of training that should be given to different pupils within the group; (3) the specific poor reading habits for which training must be given; and (4) the groupings of pupils within the class on the basis of their specific needs. This type of analysis permits the super-

¹ For example, see Table VIII, Chap. IV, p. 73.

² The data were yielded by The Stone Narrative Reading Tests, The Junior High School Examination. The chart is Fig. 2, p. 8, in *New Features of the Stone Narrative Reading Tests*, Public School Publishing Company, Bloomington, Ill.

TABLE XXVI.—A DIAGNOSTIC CHART FOR A SEVENTH-GRADE CLASS
(Stone Narrative Reading Tests—Junior High School Test)

TYPE	CLASS MEDIAN RATE 205 COMPREHENSION 12																				No. of errors TYPE			Comp.	Rate	
	III	III	I	II	II	I	III	II	II	III	I	I	II	III	III	I	I	II	I	II	III					
QUESTION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20						
William	*	*	*	*	*	*	*	*	*	*	*	*	c	*	*	*	*	*	*	*	*	1		19	167	
Stanley	*	*	*	e	a	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	2		18	273	
John	*	e	*	e	*	*	*	*	b	*	*	*	*	*	*	*	*	*	*	*	*	2	1	17	529	
Rix	a	d	*	*	*	*	*	*	d	*	*	*	*	*	*	*	*	*	*	*	*	1	2	17	265	
Doris	*	*	*	e	*	*	*	*	*	b	*	*	c	*	*	*	*	c	*	*	*	2	2	16	207	
Ruth	*	a	*	e	c	*	*	e	*	*	*	c	*	*	*	*	*	*	*	*	*	1	2	2	15	190
Margaret	*	c	*	e	d	*	*	*	*	b	e	b	*	*	*	*	*	*	*	*	*	1	3	2	14	310
Mary	a	*	*	e	b	*	*	*	*	e	c	*	*	*	*	*	c	*	*	*	*	1	3	2	14	269
Alvin	*	a	*	e	d	*	*	*	d	d	*	*	c	b	*	*	*	*	*	*	*	2	3	2	13	497
Rodger	a	*	*	e	b	*	*	*	e	*	a	e	*	*	*	*	c	*	*	e	*	3	3	2	12	357
Harry	*	a	*	b	b	a	*	*	b	e	d	*	*	*	*	*	*	*	*	a	*	6	2	12	193	
Mary B.	*	c	*	e	*	c	*	*	*	e	*	c	d	e	*	*	c	*	*	*	*	3	4	1	12	152
Mary E.	e	a	*	e	b	*	e	*	e	*	a	*	c	*	*	*	*	b	*	*	*	3	3	3	11	186
Julia	c	c	*	e	*	*	*	*	b	b	h	e	e	*	*	*	c	a	*	*	*	3	3	4	10	265
John	*	c	d	d	a	c	*	*	*	*	e	b	c	*	*	*	*	b	c	c	5	4	2	9	200	
Wilbur	e	c	*	e	b	*	*	*	e	d	*	*	b	e	*	c	a	*	b	*	2	5	4	9	145	
Lawrence	*	c	*	e	b	d	e	*	d	*	*	*	*	*	*	d	a	b	a	a	4	5	2	9	145	
William P.	c	e	a	c	c	e	*	*	a	a	a	*	e	*	*	a	*	a	*	*	2	6	4	8	167	
John S.	*	a	*	e	b	c	*	c	*	e	*	c	c	e	e	b	a	e	*	*	4	5	4	7	145	
Mary B. W.	e	c	*	e	d	e	*	c	d	a	*	*	b	*	d	c	*	d	a	d	3	6	5	6	205	
Dorothy	e	e	*	d	c	c	*	d	b	d	a	b	d	e	e	*	c	e	*	d	5	6	5	4	241	
TYPE	III	III	I	II	II	I	III	II	II	III	I	I	II	III	III	I	I	II	I	II	I	45	74	49		
Total Errors	9	16	2	19	15	8	2	8	12	10	8	10	9	5	3	5	9	8	5	5						
Incorrect responses on each question	3a 2c 4e	5a 7c 1d	1a 1d	1b 1c	2a 7b	1a 1d	2c	2b 3b	1a 3b	2a 1c	4a 6c	2b 5e	1d 2e	1a 1b	3a 6c	2a 3b	3a 1b	1a 1c	1e 2d	2e 1d	1c 2e	1d 2e	1e 2d			
Type Errors												I				II				III						
Total No. of Errors												45				74				49						
Possible Errors												147				147				126						
%												30.6				50.3				38.8						

Note: The * indicates a correct response. The letter indicates which of the four possible wrong answers was checked by the child.

visor to give the classroom teacher specific suggestions and to provide specific procedures and teaching materials for carrying out the suggestions.

This brings the discussion to the supervisor's use of tests in evaluating teaching success. No instruction program or teaching

method should be allowed to continue in use indefinitely unless it has been shown to be effective. This is true of so-called remedial instructional procedures as well as of methods devised for general use in the schools. It is possible for the supervisory staff to work out experimentally a method of teaching a given subject and to demonstrate that the method produces results. Unless there is a carefully controlled experimental comparison of the new method with other methods and a careful check as to the cost of the different methods, evidence is not available to show (1) whether or not the new method is more efficient than other methods and (2) whether or not it is economical. Whether or not the new method is adopted for general use in the school system will depend on its cost and its results; if the cost increase, if any, is considerably less than the efficiency increase, the new method may be adopted, provided funds are available to meet the increase in costs. Aspects of supervisory work in connection with teaching methods, in terms of test data, are shown in Table VI (Chapter IV, page 70) and in problems 8 and 9 (Chapter IV, pages 79-80.)

The supervisory staff must frequently select measuring instruments for use in the school system. Usually, this should be done on the basis of experimental tryout in the system itself. Data yielded by such a tryout are given in Table V (Chapter IV, page 69) and in Problem 11 (Chapter IV, page 81). Frequently, the supervisory staff must develop standardized tests. The data in Problem 7 (Chapter IV, pages 78-79) show that such tests can be developed within a school system successfully. Also, supervisors must think in terms of guiding pupils in their academic work, even though the contact work is frequently done by classroom teachers. The supervisors direct the guidance program, unless the school system has a child-guidance clinic, and they must provide the materials needed for guidance work. Two examples of research involving measurements partially are reported in Problem 12 (Chapter IV, pages 81-82) and in Table I (Chapter IV, page 62).

It has been implied in the preceding discussion that much of the supervisors' work is with classroom teachers, particularly in helping them to improve their efficiency. This in-service training, together with research, does occupy the greater part of supervisory time and attention. In order to work effectively with classroom teachers, supervisors must know their strengths and weaknesses. They can guide teachers' study and assist in

their teaching preparations when they know the teachers' abilities and needs. When supervisors understand teachers, and teachers understand supervisors, and both groups work together, both make greater and more effective efforts in the interests of the pupils.

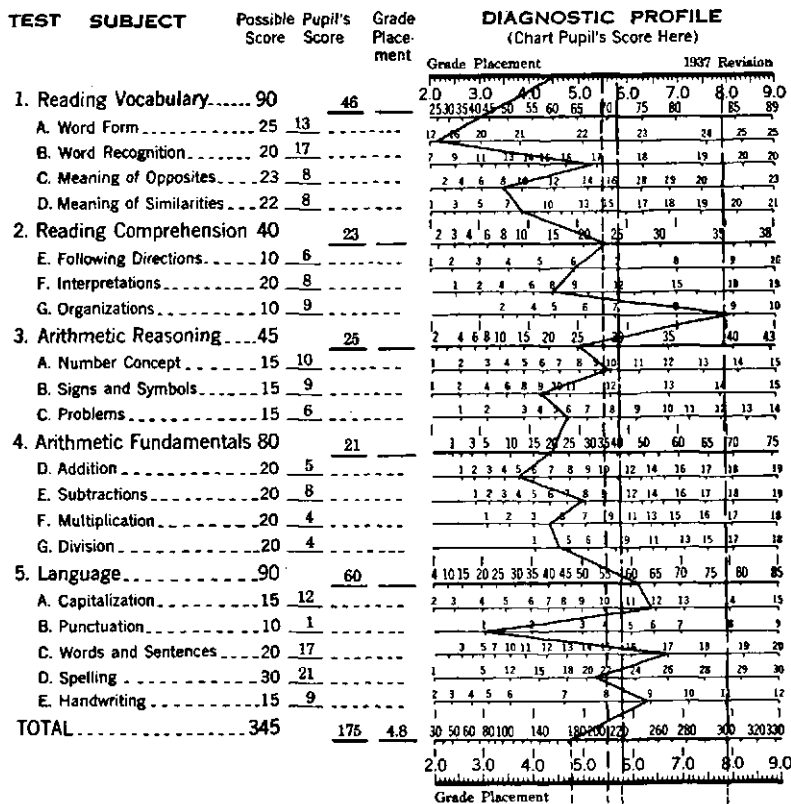
Two rather uncommon aspects of supervisory work should be mentioned. Just as it is possible to evaluate the efficiency of different teaching methods through a controlled experiment, so it is possible to set up a controlled experiment to determine the worth of different textbooks. All that is necessary is to establish balanced groups in several schools, to use a single teaching method and to have the same teachers teach each book to at least one class group, and to use identical end tests. If this is done, the better book from the standpoint of pupil achievement can be selected, since the variables have in so far as possible been controlled. Similarly, curricular materials can be checked as to their difficulty to pupils, their contribution to pupils' knowledge subsequently, and the like. These research aspects of supervision are specialized types of work, more or less unrelated directly to the general instructional program.

It should be evident from the preceding discussion that tests play a large part in supervisory work and that findings yielded by measuring instruments are as important for good supervision as they are for good school administration.

Uses in Classroom Teaching.—The important function of the classroom teacher is to guide pupil learning, bringing to the pupils what they should learn and then making certain that they learn with a minimum of error and with maximum economy of time and effort. In order to do these things, the teacher must know as much as possible about her pupils as soon as possible after they come to her for instruction. Tests, then, function to make the teacher conscious of the individual needs of her pupils and of their strengths and weaknesses. Table XXVII shows the profile of an individual pupil, indicating his subject achievements, his life and mental ages, and his actual grade placement at the beginning of a semester. This information is highly valuable to his classroom teacher, particularly since it is available when she begins to teach him.

The profile in Table XXVII permits the teacher to study the child in terms of his successes and weaknesses against certain

TABLE XXVII.—PROFILE CHART FOR A PUPIL IN GRADE 5 H¹



Achievement test grade placement -----
 Actual grade placement -----
 Normal grade placement for actual life age -----
 Normal grade placement for mental age (Stanford-Binet) -----

¹ This is the profile chart of the *Progressive Achievement Tests—Elementary Battery*—California Test Bureau, Los Angeles.

norms, his grade placement, and his mental and life ages. Studying similar profiles for the entire class group would indicate to her the specific things that need stress in teaching the particular class group. She can then check this with her supervisor and request such assistance as she deems necessary in her situation.

It is possible for the classroom teacher to translate test data into a different form in studying her individual pupils. The method is shown in Fig. 12. Here the pupil's profile is entered on the same chart as the class median profile, permitting the

teacher to grasp readily the pupil's strengths and weaknesses in terms of group achievements. This type of evidence, as well as that in Table XXVII, are valuable for the teacher's use in talking both with the pupil and with his parents concerning his schoolwork.

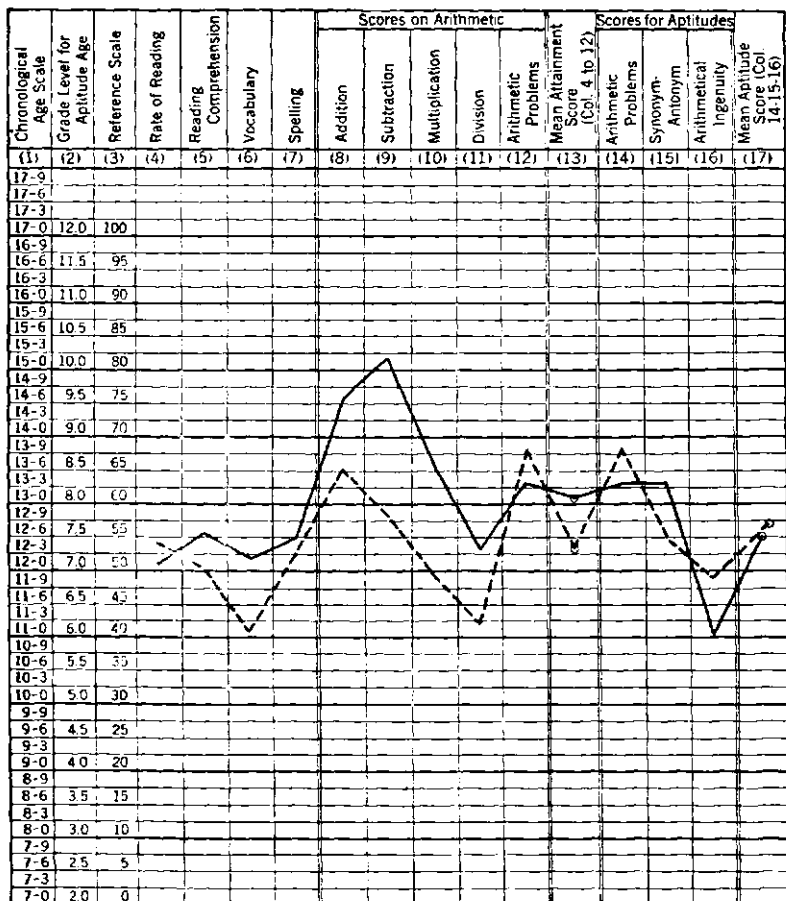


FIG. 12.—Comparison of a pupil's test scores with the class scores on the Public School Correlated Attainment Scales, Form C.¹ (Seventh-grade class tested in February. The scale age for February is 12 years, 7 months, or a score of 56%.)

Solid line graph represents the test scores of a pupil of median aptitude. For the attainment tests, it shows that this pupil is above the work of his class in all tests except in "Rate of Reading" and "Arithmetic Problems."

Broken line graph represents the class means.

¹Published by the Public School Publishing Company, Bloomington, Ill. The figure is from page 5, Test Bulletin No. 1, *The Testing Program*.

This naturally leads to a consideration of the teacher's guidance work with pupils. Guidance is (1) for the future and (2) for the immediate present. Teachers must advise parents with respect to the type of work that their children should undertake in secondary schools, and they should attempt to give some form of vocational guidance as well, preferably in general terms. The particular guidance function of the elementary-school teacher is for immediate purposes. This has to do with the assistance given directly to pupils, based on an accurate knowledge of what pupils have done in their past schoolwork and of their mental ability. Relatively little has been done directly in guidance in the elementary school.

Test data have numerous other uses in elementary-school teaching. Test results are useful for grouping, in the relatively few situations in which grouping is desirable. They form one basis for planning types of work and for establishing the speed of work to be done by the class group. They help to insure more accurate school marks, and they help also to keep the pupils working somewhere near their actual ability. Finally, they help the teacher to judge her own teaching successes and failures in terms of the achievements of her pupils. Such a knowledge permits her to use the facilities provided by the school administration for her assistance in improving her teaching and particularly that of the supervisory staff and of the research department of the school system.

Summary.—This chapter has attempted to present some specific uses of measuring instruments in educational work in elementary schools. Reference has been made to the materials presented earlier in this volume. The sum total should make clear that tests definitely have a place in the education of children in elementary schools and that there are uses of which administrators, supervisors, and classroom teachers should avail themselves. It should be noted again, however, that tests do not by any means constitute the only source of information available to educational workers and that measurement is not the whole of any educator's business. Tests should be used extensively, whenever and wherever they are of value, but not exclusively in the process of guiding the instruction of pupils in the elementary school.

Readings for Further Study

1. A. S. BARR and W. H. BURTON, *The Supervision of Instruction*. Chap. IX. New York: D. Appleton-Century Company, Inc., 1926.
2. T. H. BRIGGS, *Improving Instruction*. Chap. XVII. New York: The Macmillan Company, 1938.
3. L. J. BRUECKNER and E. O. MELBY, *Diagnostic and Remedial Teaching*. Chaps. IV, V, VII-XII, XIV.
4. A. J. JONES, *Principles of Guidance*. Chap. IX. New York: McGraw-Hill Book Company, Inc., 1934.
5. J. M. LEE, *A Guide to Measurement in Secondary Schools*. Chaps. III, V, VII-IX.
6. W. A. McCALL, *How to Measure in Education*. Chaps. II-VI.
7. G. M. WILSON and K. J. HOKE, *How to Measure* (rev. ed.). Chap. XXVI.
8. C. WOODY and P. V. SANGREN, *Administration of the Testing Program*. Chaps. VII-IX.

Questions and Exercises

1. You are a beginning teacher in grade 6 in an elementary-school system, without previous teaching experience. In so far as tests are concerned, what should be your procedure? By whom should you be guided?
2. Assume that you have taught for seven years in grade 6. You have a group of pupils numbering 33, 12 with I. Q. below 90, 6 with I. Q.s between 90 and 110, and the remainder with I. Q.s above 110. What achievement tests should you give this group in attempting to plan for your semester of teaching?
3. What measuring procedure should a teacher follow in teaching an upper grade social science activity unit? (NOTE.—Standardized tests are not yet available for measuring the results of activity teaching.)
4. A superintendent is thinking of introducing a semester of reading-readiness instruction in grade 1 L. What should be his procedure in so far as measurements is concerned?
5. Indicate the overlapping uses of measurements, both achievement and mental tests, among administrators, supervisors, and classroom teachers. In what ways do these different persons use test data similarly? Differently?
6. Write a news story reporting the results of an achievement-test survey of a small school system, in which reading and arithmetic are superior and English grammar and spelling are retarded. In your story, indicate what is proposed in remedying the situation.
7. What training in educational measurements should educators in elementary schools possess?

APPENDIX A

PUBLISHERS OF STANDARDIZED TESTS

The names and addresses of the companies that publish educational tests are given in the following list:

- American Child Health Association, New York City
- A. S. Barnes and Co., New York City
- Association Press, New York City
- Bruce Publishing Co., Milwaukee, Wis.
- Bureau of Educational Research and Service, University of Iowa, Iowa City
- C. A. Gregory Co., 345 Calhoun St., Cincinnati, Ohio
- California Test Bureau, 3636 Beverly Boulevard, Los Angeles
- Catholic Education Press, 1326 Quincy Street, N. E., Washington, D. C.
- Center for Psychological Service, Washington, D. C.
- Charles Scribner's Sons, New York City
- C. H. Stoelting Co., New York City
- Cooperative Test Service, New York City
- D. C. Heath and Co., Boston, Mass.
- Department of Synagogue and School Extension, Union of Ameneah Hebrew Congregations, Cincinnati, Ohio
- Educational Test Bureau, Minneapolis, Minn.
- Ginn and Company, Boston, Mass.
- Gregg Publishing Co., New York City
- Harter School Supply Co., Cleveland, Ohio
- Harvard University Press, Cambridge, Mass.
- Houghton Mifflin Company, Boston, Mass.
- John C. Winston Company, Philadelphia, Pa.
- Kansas State Teachers College, Emporia
- Lafayette Printing Co., Lafayette, Ind.
- Lyons and Carnahan, Chicago, Ill.
- Manual Arts Press, Peoria, Ill.
- McKnight and McKnight, Bloomington, Ill.
- Palmer Co., Boston, Mass.
- Public School Publishing Company, Bloomington, Ill.
- Rural Education Dept., State College, Pa.
- Scott, Foresman & Company, Chicago, Ill.
- Smith, Hammond and Co., Atlanta, Ga.
- Southwestern Publishing Co., Cincinnati, Ohio
- Stanford University Press, Stanford University, Calif.
- Bureau of Publications, Teachers College, Columbia University, New York City
- University of Chicago Press, Chicago
- Warwick and York, Baltimore, Md.
- World Book Company, Yonkers-on-Hudson, N. Y.

322 MEASUREMENTS IN THE ELEMENTARY SCHOOL

Form for the computation of the E. C.

 M_x : σ_x : N_x : M_y : σ_y : N_y :

$$E. C. = \frac{\quad}{2.78 \quad x \sqrt{\frac{\quad}{2} + \frac{\quad}{2}}}$$

$$= \frac{\quad}{2.78 \quad x \sqrt{\quad + \quad}}$$

$$= \frac{\quad}{2.78 \quad x \sqrt{\quad + \quad}}$$

$$= \frac{\quad}{2.78 \quad x \sqrt{\quad}}$$

$$= \frac{\quad}{2.78 \quad x}$$

$$= \frac{\quad}{\quad}$$

E. C. =

Answer

APPENDIX C

SQUARES AND SQUARE ROOTS OF NUMBERS FROM 1 TO 1000

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
1	1	1.000	41	16 81	6.403
2	4	1.414	42	17 64	6.481
3	9	1.732	43	18 49	6.557
4	16	2.000	44	19 36	6.633
5	25	2.236	45	20 25	6.708
6	36	2.449	46	21 16	6.782
7	49	2.646	47	22 09	6.856
8	64	2.828	48	23 04	6.928
9	81	3.000	49	24 01	7.000
10	1 00	3.162	50	25 00	7.071
11	1 21	3.317	51	26 01	7.141
12	1 44	3.464	52	27 04	7.211
13	1 69	3.606	53	28 09	7.280
14	1 96	3.742	54	29 16	7.348
15	2 25	3.873	55	30 25	7.416
16	2 56	4.000	56	31 36	7.483
17	2 89	4.123	57	32 49	7.550
18	3 24	4.243	58	33 64	7.616
19	3 61	4.359	59	34 81	7.681
20	4 00	4.472	60	36 00	7.746
21	4 41	4.583	61	37 21	7.810
22	4 84	4.690	62	38 44	7.874
23	5 29	4.796	63	39 69	7.937
24	5 76	4.899	64	40 96	8.000
25	6 25	5.000	65	42 25	8.062
26	6 76	5.099	66	43 56	8.124
27	7 29	5.196	67	44 89	8.185
28	7 84	5.292	68	46 24	8.246
29	8 41	5.385	69	47 61	8.307
30	9 00	5.477	70	49 00	8.367
31	9 61	5.568	71	50 41	8.426
32	10 24	5.657	72	51 84	8.485
33	10 89	5.745	73	53 29	8.544
34	11 56	5.831	74	54 76	8.602
35	12 25	5.916	75	56 25	8.660
36	12 96	6.000	76	57 76	8.718
37	13 69	6.083	77	59 29	8.775
38	14 44	6.164	78	60 84	8.832
39	15 21	6.245	79	62 41	8.888
40	16 00	6.325	80	64 00	8.944

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
81	65 61	9.000	126	1 58 76	11.225
82	67 24	9.055	127	1 61 29	11.269
83	68 89	9.110	128	1 63 84	11.314
84	70 56	9.165	129	1 66 41	11.358
85	72 25	9.220	130	1 69 00	11.402
86	73 96	9.274	131	1 71 61	11.446
87	75 69	9.327	132	1 74 24	11.489
88	77 44	9.381	133	1 76 89	11.533
89	79 21	9.434	134	1 79 56	11.576
90	81 00	9.487	135	1 82 25	11.619
91	82 81	9.539	136	1 84 96	11.662
92	84 64	9.592	137	1 87 69	11.705
93	86 49	9.644	138	1 90 44	11.747
94	88 36	9.695	139	1 93 21	11.790
95	90 25	9.747	140	1 96 00	11.832
96	92 16	9.798	141	1 98 81	11.874
97	94 09	9.849	142	2 01 64	11.916
98	96 04	9.899	143	2 04 49	11.958
99	98 01	9.950	144	2 07 36	12.000
100	1 00 00	10.000	145	2 10 25	12.042
101	1 02 01	10.050	146	2 13 16	12.083
102	1 04 04	10.100	147	2 16 09	12.124
103	1 06 09	10.149	148	2 19 04	12.166
104	1 08 16	10.198	149	2 22 01	12.207
105	1 10 25	10.247	150	2 25 00	12.247
106	1 12 36	10.296	151	2 28 01	12.288
107	1 14 49	10.344	152	2 31 04	12.329
108	1 16 64	10.392	153	2 34 09	12.369
109	1 18 81	10.440	154	2 37 16	12.410
110	1 21 00	10.488	155	2 40 25	12.450
111	1 23 21	10.536	156	2 43 36	12.490
112	1 25 44	10.583	157	2 46 49	12.530
113	1 27 69	10.630	158	2 49 64	12.570
114	1 29 96	10.677	159	2 52 81	12.610
115	1 32 25	10.724	160	2 56 00	12.649
116	1 34 56	10.770	161	2 59 21	12.689
117	1 36 89	10.817	162	2 62 44	12.728
118	1 39 24	10.863	163	2 65 69	12.767
119	1 41 61	10.909	164	2 68 96	12.806
120	1 44 00	10.954	165	2 72 25	12.845
121	1 46 41	11.000	166	2 75 56	12.884
122	1 48 84	11.045	167	2 78 89	12.923
123	1 51 20	11.091	168	2 82 24	12.961
124	1 53 76	11.136	169	2 85 61	13.000
125	1 56 25	11.180	170	2 89 00	13.038

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
171	2 92 41	13.077	216	4 66 56	14.697
172	2 95 84	13.115	217	4 70 89	14.731
173	2 99 29	13.153	218	4 75 24	14.765
174	3 02 76	13.191	219	4 79 61	14.799
175	3 06 25	13.229	220	4 84 00	14.832
176	3 09 76	13.266	221	4 88 41	14.866
177	3 13 29	13.304	222	4 92 84	14.900
178	3 16 84	13.342	223	4 97 29	14.933
179	3 20 41	13.379	224	5 01 76	14.967
180	3 24 00	13.416	225	5 06 25	15.000
181	3 27 61	13.454	226	5 10 76	15.033
182	3 31 24	13.491	227	5 15 29	15.067
183	3 34 89	13.528	228	5 19 84	15.100
184	3 38 56	13.565	229	5 24 41	15.133
185	3 42 25	13.601	230	5 29 00	15.166
186	3 45 96	13.638	231	5 33 61	15.199
187	3 49 69	13.675	232	5 38 24	15.232
188	3 53 44	13.711	233	5 42 89	15.264
189	3 57 21	13.748	234	5 47 56	15.297
190	3 61 00	13.784	235	5 52 25	15.330
191	3 64 81	13.820	236	5 56 96	15.362
192	3 68 64	13.856	237	5 61 69	15.395
193	3 72 49	13.892	238	5 66 44	15.427
194	3 76 36	13.928	239	5 71 21	15.460
195	3 80 25	13.964	240	5 76 00	15.492
196	3 84 16	14.000	241	5 80 81	15.524
197	3 88 09	14.036	242	5 85 64	15.556
198	3 92 04	14.071	243	5 90 49	15.588
199	3 96 01	14.107	244	5 95 36	15.620
200	4 00 00	14.142	245	6 00 25	15.652
201	4 04 01	14.177	246	6 05 16	15.684
202	4 08 04	14.213	247	6 10 09	15.716
203	4 12 09	14.248	248	6 15 04	15.748
204	4 16 16	14.283	249	6 20 01	15.780
205	4 20 25	14.318	250	6 25 00	15.811
206	4 24 36	14.353	251	6 30 01	15.843
207	4 28 49	14.387	252	6 35 04	15.875
208	4 32 64	14.422	253	6 40 09	15.906
209	4 36 81	14.457	254	6 45 16	15.937
210	4 41 00	14.491	255	6 50 25	15.969
211	4 45 21	14.526	256	6 55 36	16.000
212	4 49 44	14.560	257	6 60 49	16.031
213	4 53 69	14.595	258	6 65 64	16.062
214	4 57 96	14.629	259	6 70 81	16.093
215	4 62 25	14.663	260	6 76 00	16.125

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
261	6 81 21	16.155	306	9 36 36	17.493
262	6 86 44	16.186	307	9 42 49	17.521
263	6 91 69	16.217	308	9 48 64	17.550
264	6 96 96	16.248	309	9 54 81	17.578
265	7 02 25	16.279	310	9 61 00	17.607
266	7 07 56	16.310	311	9 67 21	17.635
267	7 12 89	16.340	312	9 73 44	17.664
268	7 18 24	16.371	313	9 79 69	17.692
269	7 23 61	16.401	314	9 85 96	17.720
270	7 29 00	16.432	315	9 92 25	17.748
271	7 34 41	16.462	316	9 98 56	17.776
272	7 39 84	16.492	317	10 04 89	17.804
273	7 45 29	16.523	318	10 11 24	17.833
274	7 50 76	16.553	319	10 17 61	17.861
275	7 56 25	16.583	320	10 24 00	17.889
276	7 61 76	16.613	321	10 30 41	17.916
277	7 67 29	16.643	322	10 36 84	17.944
278	7 72 84	16.673	323	10 43 29	17.972
279	7 78 41	16.703	324	10 49 76	18.000
280	7 84 00	16.733	325	10 56 25	18.028
281	7 89 61	16.763	326	10 62 76	18.055
282	7 95 24	16.793	327	10 69 29	18.083
283	8 00 89	16.823	328	10 75 84	18.111
284	8 06 56	16.852	329	10 82 41	18.138
285	8 12 25	16.882	330	10 89 00	18.166
286	8 17 96	16.912	331	10 95 61	18.193
287	8 23 69	16.941	332	11 02 24	18.221
288	8 29 44	16.971	333	11 08 89	18.248
289	8 35 21	17.000	334	11 15 56	18.276
290	8 41 00	17.029	335	11 22 25	18.303
291	8 46 81	17.059	336	11 28 96	18.330
292	8 52 64	17.088	337	11 35 69	18.358
293	8 58 49	17.117	338	11 42 44	18.385
294	8 64 36	17.146	339	11 49 21	18.412
295	8 70 25	17.176	340	11 56 00	18.439
296	8 76 16	17.205	341	11 62 81	18.466
297	8 82 09	17.234	342	11 69 64	18.493
298	8 88 04	17.263	343	11 76 49	18.520
299	8 94 01	17.292	344	11 83 36	18.547
300	9 00 00	17.321	345	11 90 25	18.574
301	9 06 01	17.349	346	11 97 16	18.601
302	9 12 04	17.378	347	12 04 09	18.628
303	9 18 09	17.407	348	12 11 04	18.655
304	9 24 16	17.436	349	12 18 01	18.682
305	9 30 25	17.464	350	12 25 00	18.708

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
351	12 32 01	18.735	396	15 68 16	19.900
352	12 39 04	18.762	397	15 76 09	19.925
353	12 46 09	18.788	398	15 84 04	19.950
354	12 53 16	18.815	399	15 92 01	19.975
355	12 60 25	18.841	400	16 00 00	20.000
356	12 67 36	18.868	401	16 08 01	20.025
357	12 74 49	18.894	402	16 16 04	20.050
358	12 81 64	18.921	403	16 24 09	20.075
359	12 88 81	18.947	404	16 32 16	20.100
360	12 96 00	18.974	405	16 40 25	20.125
361	13 03 21	19.000	406	16 48 36	20.149
362	13 10 44	19.026	407	16 56 49	20.174
363	13 17 69	19.053	408	16 64 64	20.199
364	13 24 96	19.079	409	16 72 81	20.224
365	13 32 25	19.105	410	16 81 00	20.248
366	13 39 56	19.131	411	16 89 21	20.273
367	13 46 89	19.157	412	16 97 44	20.298
368	13 54 24	19.183	413	17 05 69	20.322
369	13 61 61	19.209	414	17 13 96	20.347
370	13 69 00	19.235	415	17 22 25	20.372
371	13 76 41	19.261	416	17 30 56	20.396
372	13 83 84	19.287	417	17 38 89	20.421
373	13 91 29	19.313	418	17 47 24	20.445
374	13 98 76	19.339	419	17 55 61	20.469
375	14 06 25	19.365	420	17 64 00	20.494
376	14 13 76	19.391	421	17 72 41	20.518
377	14 21 29	19.416	422	17 80 84	20.543
378	14 28 84	19.442	423	17 89 29	20.567
379	14 36 41	19.468	424	17 97 76	20.591
380	14 44 00	19.494	425	18 06 25	20.616
381	14 51 61	19.519	426	18 14 76	20.640
382	14 59 24	19.545	427	18 23 29	20.664
383	14 66 89	19.570	428	18 31 84	20.688
384	14 74 56	19.596	429	18 40 41	20.712
385	14 82 25	19.621	430	18 49 00	20.736
386	14 89 96	19.647	431	18 57 61	20.761
387	14 97 69	19.672	432	18 66 24	20.785
388	15 05 44	19.698	433	18 74 89	20.809
389	15 13 21	19.723	434	18 83 56	20.833
390	15 21 00	19.748	435	18 92 25	20.857
391	15 28 81	19.774	436	19 00 96	20.881
392	15 36 64	19.799	437	19 09 69	20.905
393	15 44 49	19.824	438	19 18 44	20.928
394	15 52 36	19.849	439	19 27 21	20.952
395	15 60 25	19.875	440	19 36 00	20.976

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
441	19 44 81	21.000	486	23 61 96	22.045
442	19 53 64	21.024	487	23 71 69	22.068
443	19 62 49	21.048	488	23 81 44	22.091
444	19 71 36	21.071	489	23 91 21	22.113
445	19 80 25	21.095	490	24 01 00	22.136
446	19 89 16	21.119	491	24 10 81	22.159
447	19 98 09	21.142	492	24 20 64	22.181
448	20 07 04	21.166	493	24 30 49	22.204
449	20 16 01	21.190	494	24 40 36	22.226
450	20 25 00	21.213	495	24 50 25	22.249
451	20 34 01	21.237	496	24 60 16	22.271
452	20 43 04	21.260	497	24 70 09	22.293
453	20 52 09	21.284	498	24 80 04	22.316
454	20 61 16	21.307	499	24 90 01	22.338
455	20 70 25	21.331	500	25 00 00	22.361
456	20 79 36	21.354	501	25 10 01	22.383
457	20 88 49	21.378	502	25 20 04	22.405
458	20 97 64	21.401	503	25 30 09	22.428
459	21 06 81	21.424	504	25 40 16	22.450
460	21 16 00	21.448	505	25 50 25	22.472
461	21 25 21	21.471	506	25 60 36	22.494
462	21 34 44	21.494	507	25 70 49	22.517
463	21 43 69	21.517	508	25 80 64	22.539
464	21 52 96	21.541	509	25 90 81	22.561
465	21 62 25	21.564	510	26 01 00	22.583
466	21 71 56	21.587	511	26 11 21	22.605
467	21 80 89	21.610	512	26 21 44	22.627
468	21 90 24	21.633	513	26 31 69	22.650
469	21 99 61	21.656	514	26 41 96	22.672
470	22 09 00	21.679	515	26 52 25	22.694
471	22 18 41	21.703	516	26 62 56	22.716
472	22 27 84	21.726	517	26 72 89	22.738
473	22 37 29	21.749	518	26 83 24	22.760
474	22 46 76	21.772	519	26 93 61	22.782
475	22 56 25	21.794	520	27 04 00	22.804
476	22 65 76	21.817	521	27 14 41	22.825
477	22 75 29	21.840	522	27 24 84	22.847
478	22 84 84	21.863	523	27 35 29	22.869
479	22 94 41	21.886	524	27 45 76	22.891
480	23 04 00	21.909	525	27 56 25	22.913
481	23 13 61	21.932	526	27 66 76	22.935
482	23 23 24	21.954	527	27 77 29	22.956
483	23 32 89	21.977	528	27 87 84	22.978
484	23 42 56	22.000	529	27 98 41	23.000
485	23 52 25	22.023	530	28 09 00	23.022

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
531	28 19 61	23.043	581	33 75 61	24.104
532	28 30 24	23.065	582	33 87 24	24.125
533	28 40 89	23.087	583	33 98 89	24.145
534	28 51 56	23.108	584	34 10 56	24.166
535	28 62 25	23.130	585	34 22 25	24.187
536	28 72 96	23.152	586	34 33 96	24.207
537	28 83 69	23.173	587	34 45 69	24.228
538	28 94 44	23.195	588	34 57 44	24.249
539	29 05 21	23.216	589	34 69 21	24.269
540	29 16 00	23.238	590	34 81 00	24.290
541	29 26 81	23.259	591	34 92 81	24.310
542	29 37 64	23.281	592	35 04 64	24.331
543	29 48 49	23.302	593	35 16 49	24.352
544	29 59 36	23.324	594	35 28 36	24.372
545	29 70 25	23.345	595	35 40 25	24.393
546	29 81 16	23.367	596	35 52 16	24.413
547	29 92 09	23.388	597	35 64 09	24.434
548	30 03 04	23.409	598	35 76 04	24.454
549	30 14 01	23.431	599	35 88 01	24.474
550	30 25 00	23.452	600	36 00 00	24.495
551	30 36 01	23.473	601	36 12 01	24.515
552	30 47 04	23.495	602	36 24 04	24.536
553	30 58 09	23.516	603	36 36 09	24.556
554	30 69 16	23.537	604	36 48 16	24.576
555	30 80 25	23.558	605	36 60 25	24.597
556	30 91 36	23.580	606	36 72 36	24.617
557	31 02 49	23.601	607	36 84 49	24.637
558	31 13 64	23.622	608	36 96 64	24.658
559	31 24 81	23.643	609	37 08 81	24.678
560	31 36 00	23.664	610	37 21 00	24.698
561	31 47 21	23.685	611	37 33 21	24.718
562	31 58 44	23.707	612	37 45 44	24.739
563	31 69 69	23.728	613	37 57 69	24.759
564	31 80 96	23.749	614	37 69 96	24.779
565	31 92 25	23.770	615	37 82 25	24.799
566	32 03 56	23.791	616	37 94 56	24.819
567	32 14 89	23.812	617	38 06 89	24.839
568	32 26 24	23.833	618	38 19 24	24.860
569	32 37 61	23.854	619	38 31 61	24.880
570	32 49 00	23.875	620	38 44 00	24.900
571	32 60 41	23.896	621	38 56 41	24.920
572	32 71 84	23.917	622	38 68 84	24.940
573	32 83 29	23.937	623	38 81 29	24.960
574	32 94 76	23.958	624	38 93 76	24.980
575	33 06 25	23.979	625	39 06 25	25.000
576	33 17 76	24.000	626	39 18 76	25.020
577	33 29 29	24.021	627	39 31 29	25.040
578	33 40 84	24.042	628	39 43 84	25.059
579	33 52 41	24.062	629	39 56 41	25.080
580	33 64 00	24.083	630	39 69 00	25.100

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
631	39 81 61	25.120	681	46 37 61	26.096
632	39 94 24	25.140	682	46 51 24	26.115
633	40 06 89	25.159	683	46 64 89	26.134
634	40 19 56	25.179	684	46 78 56	26.153
635	40 32 25	25.199	685	46 92 25	26.173
636	40 44 96	25.219	686	47 05 96	26.192
637	40 57 69	25.239	687	47 19 69	26.211
638	40 70 44	25.259	688	47 33 44	26.230
639	40 83 21	25.278	689	47 47 21	26.249
640	40 96 00	25.298	690	47 61 00	26.268
641	41 08 81	25.318	691	47 74 81	26.287
642	41 21 64	25.338	692	47 88 64	26.306
643	41 34 49	25.357	693	48 02 49	26.325
644	41 47 36	25.377	694	48 16 36	26.344
645	41 60 25	25.397	695	48 30 25	26.363
646	41 73 16	25.417	696	48 44 16	26.382
647	41 86 09	25.436	697	48 58 09	26.401
648	41 99 04	25.456	698	48 72 04	26.420
649	42 12 01	25.475	699	48 86 01	26.439
650	42 25 00	25.495	700	49 00 00	26.458
651	42 38 01	25.515	701	49 14 01	26.476
652	42 51 04	25.534	702	49 28 04	26.495
653	42 64 09	25.554	703	49 42 09	26.514
654	42 77 16	25.573	704	49 56 16	26.533
655	42 90 25	25.593	705	49 70 25	26.552
656	43 03 36	25.612	706	49 84 36	26.571
657	43 16 49	25.632	707	49 98 49	26.589
658	43 29 64	25.652	708	50 12 64	26.608
659	43 42 81	25.671	709	50 26 81	26.627
660	43 56 00	25.690	710	50 41 00	26.646
661	43 69 21	25.710	711	50 55 21	26.665
662	43 82 44	25.729	712	50 69 44	26.683
663	43 95 69	25.749	713	50 83 69	26.702
664	44 08 96	25.768	714	50 97 96	26.721
665	44 22 25	25.788	715	51 12 25	26.739
666	44 35 56	25.807	716	51 26 56	26.758
667	44 48 89	25.826	717	51 40 89	26.777
668	44 62 24	25.846	718	51 55 24	26.796
669	44 75 61	25.865	719	51 69 61	26.814
670	44 89 00	25.884	720	51 84 00	26.833
671	45 02 41	25.904	721	51 98 41	26.851
672	45 15 84	25.923	722	52 12 84	26.870
673	45 29 29	25.942	723	52 27 29	26.889
674	45 42 76	25.962	724	52 41 76	26.907
675	45 56 25	25.981	725	52 56 25	26.926
676	45 69 76	26.000	726	52 70 76	26.944
677	45 83 29	26.019	727	52 85 29	26.963
678	45 96 84	26.038	728	52 99 84	26.981
679	46 10 41	26.058	729	53 14 41	27.000
680	46 24 00	26.077	730	53 29 00	27.019

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
731	53 43 61	27.037	776	60 21 76	27.857
732	53 58 24	27.055	777	60 37 29	27.875
733	53 72 89	27.074	778	60 52 84	27.893
734	53 87 56	27.092	779	60 68 41	27.911
735	54 02 25	27.111	780	60 84 00	27.928
736	54 16 96	27.129	781	60 99 61	27.946
737	54 31 69	27.148	782	61 15 24	27.964
738	54 46 44	27.166	783	61 30 89	27.982
739	54 61 21	27.185	784	61 46 56	28.000
740	54 76 00	27.203	785	61 62 25	28.018
741	54 90 81	27.221	786	61 77 96	28.036
742	55 05 64	27.240	787	61 93 69	28.054
743	55 20 49	27.258	788	62 09 44	28.071
744	55 35 36	27.276	789	62 25 21	28.089
745	55 50 25	27.295	790	62 41 00	28.107
746	55 65 16	27.313	791	62 56 81	28.125
747	55 80 09	27.331	792	62 72 64	28.142
748	55 95 04	27.350	793	62 88 49	28.160
749	56 10 01	27.368	794	63 04 36	28.178
750	56 25 00	27.386	795	63 20 25	28.196
751	56 40 01	27.404	796	63 36 16	28.213
752	56 55 04	27.423	797	63 52 09	28.231
753	56 70 09	27.441	798	63 68 04	28.249
754	56 85 16	27.459	799	63 84 01	28.267
755	57 00 25	27.477	800	64 00 00	28.284
756	57 15 36	27.495	801	64 16 01	28.302
757	57 30 49	27.514	802	64 32 04	28.320
758	57 45 64	27.532	803	64 48 09	28.337
759	57 60 81	27.550	804	64 64 16	28.355
760	57 76 00	27.568	805	64 80 25	28.373
761	57 91 21	27.586	806	64 96 36	28.390
762	58 06 44	27.604	807	65 12 49	28.408
763	58 21 69	27.622	808	65 28 64	28.425
764	58 36 96	27.641	809	65 44 81	28.443
765	58 52 25	27.659	810	65 61 00	28.460
766	58 67 56	27.677	811	65 77 21	28.478
767	58 82 89	27.695	812	65 93 44	28.496
768	58 98 24	27.713	813	66 09 69	28.513
769	59 13 61	27.731	814	66 25 96	28.531
770	59 29 00	27.749	815	66 42 25	28.548
771	59 44 41	27.767	816	66 58 56	28.566
772	59 59 84	27.785	817	66 74 89	28.583
773	59 75 29	27.803	818	66 91 24	28.601
774	59 90 76	27.821	819	67 07 61	28.618
775	60 06 25	27.839	820	67 24 00	28.636

332 MEASUREMENTS IN THE ELEMENTARY SCHOOL

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
821	67 40 41	28.653	866	74 99 56	29.428
822	67 56 84	28.671	867	75 16 89	29.445
823	67 73 29	28.688	868	75 34 24	29.462
824	67 89 76	28.705	869	75 51 61	29.479
825	68 06 25	28.723	870	75 69 00	29.496
826	68 22 76	28.740	871	75 86 41	29.513
827	68 39 29	28.758	872	76 03 84	29.530
828	68 55 84	28.775	873	76 21 29	29.547
829	68 72 41	28.792	874	76 38 76	29.563
830	68 89 00	28.810	875	76 56 25	29.580
831	69 05 61	28.827	876	76 73 76	29.597
832	69 22 24	28.844	877	76 91 29	29.614
833	69 38 89	28.862	878	77 08 84	29.631
834	69 55 56	28.879	879	77 26 41	29.648
835	69 72 25	28.896	880	77 44 00	29.665
836	69 88 96	28.914	881	77 61 61	29.682
837	70 05 69	28.931	882	77 79 24	29.698
838	70 22 44	28.948	883	77 96 89	29.715
839	70 39 21	28.965	884	78 14 56	29.732
840	70 56 00	28.983	885	78 32 25	29.749
841	70 72 81	29.000	886	78 49 96	29.766
842	70 89 64	29.017	887	78 67 69	29.783
843	71 06 49	29.034	888	78 85 44	29.799
844	71 23 36	29.052	889	79 03 21	29.816
845	71 40 25	29.069	890	79 21 00	29.833
846	71 57 16	29.086	891	79 38 81	29.850
847	71 74 09	29.103	892	79 56 64	29.866
848	71 91 04	29.120	893	79 74 49	29.883
849	72 08 01	29.138	894	79 92 36	29.900
850	72 25 00	29.155	895	80 10 25	29.917
851	72 42 01	29.172	896	80 28 16	29.933
852	72 59 04	29.189	897	80 46 09	29.950
853	72 76 09	29.206	898	80 64 04	29.967
854	72 93 16	29.223	899	80 82 01	29.983
855	73 10 25	29.240	900	81 00 00	30.000
856	73 27 36	29.257	901	81 18 01	30.017
857	73 44 49	29.275	902	81 36 04	30.033
858	73 61 64	29.292	903	81 54 09	30.050
859	73 78 81	29.309	904	81 72 16	30.067
860	73 96 00	29.326	905	81 90 25	30.083
861	74 13 21	29.343	906	82 08 36	30.100
862	74 30 44	29.360	907	82 26 49	30.116
863	74 47 69	29.377	908	82 44 64	30.133
864	74 64 96	29.394	909	82 62 81	30.150
865	74 82 25	29.411	910	82 81 00	30.156

SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM 1 TO 1000

Number	Square	Square root	Number	Square	Square root
911	82 99 21	30.183	956	91 39 36	30.919
912	83 17 44	30.199	957	91 58 49	30.935
913	83 35 69	30.216	958	91 77 64	30.952
914	83 53 96	30.232	959	91 96 81	30.968
915	83 72 25	30.249	960	92 16 00	30.984
916	83 90 56	30.265	961	92 35 21	31.000
917	84 08 89	30.282	962	92 54 44	31.016
918	84 27 24	30.299	963	92 73 69	31.032
919	84 45 61	30.315	964	92 92 96	31.048
920	84 64 00	30.332	965	93 12 25	31.064
921	84 82 41	30.348	966	93 31 56	31.081
922	85 00 84	30.364	967	93 50 89	31.097
923	85 19 29	30.381	968	93 70 24	31.113
924	85 37 76	30.397	969	93 89 61	31.129
925	85 56 25	30.414	970	94 09 00	31.145
926	85 74 76	30.430	971	94 28 41	31.161
927	85 93 29	30.447	972	94 47 84	31.177
928	86 11 84	30.463	973	94 67 29	31.193
929	86 30 41	30.480	974	94 86 76	31.209
930	86 49 00	30.496	975	95 06 25	31.225
931	86 67 61	30.512	976	95 25 76	31.241
932	86 86 24	30.529	977	95 45 29	31.257
933	87 04 89	30.545	978	95 64 84	31.273
934	87 23 56	30.561	979	95 84 41	31.289
935	87 42 25	30.578	980	96 04 00	31.305
936	87 60 96	30.594	981	96 23 61	31.321
937	87 79 69	30.610	982	96 43 24	31.337
938	87 98 44	30.627	983	96 62 89	31.353
939	88 17 21	30.643	984	96 82 56	31.369
940	88 36 00	30.659	985	97 02 25	31.385
941	88 54 81	30.676	986	97 21 96	31.401
942	88 73 64	30.692	987	97 41 69	31.417
943	88 92 49	30.708	988	97 61 44	31.432
944	89 11 36	30.725	989	97 81 21	31.448
945	89 30 25	30.741	990	98 01 00	31.464
946	89 49 16	30.757	991	98 20 81	31.480
947	89 68 09	30.773	992	98 40 64	31.496
948	89 87 04	30.790	993	98 60 49	31.512
949	90 06 01	30.806	994	98 80 36	31.528
950	90 25 00	30.822	995	99 00 25	31.544
951	90 44 01	30.838	996	99 20 16	31.559
952	90 63 04	30.854	997	99 40 09	31.575
953	90 82 09	30.871	998	99 60 04	31.591
954	91 01 16	30.887	999	99 80 01	31.607
955	91 20 25	30.903	1000	1 00 00 00	31.623

INDEX

A

- Abbott, A., 160*n*.
Abilities of Man, The, 259*n.*, 274
 Accomplishment age, 20
 Accomplishment quotient, 20-21
Adjustment Questionnaire, A Group Factor Analysis of the, 300
 Advantages of tests (*see* Objective tests, advantages of; Standardized tests, advantages of)
 Aesthetics, defined, 268-269
 Algebra tests, 91-93
 Allen, R. D., 169*n*.
 Anderson, R. G., 286*n.*, 287*n*.
 Anderson, W. A., 168*n*.
Aptitude Testing, 260*n.*, 300
 Arithmetic, teaching and testing of 207-209, 222-223
 Arithmetic mean, 30-32, 55-57
 computation form for, 320
 Arithmetic tests, 207-223
 diagnostic, 208-218
 reasoning, 216-219
 survey, 218-222
 uses of, 222-223
 Armstrong, C., 180*n.*, 218*n.*, 219*n*.
 Art teaching, purpose of, 236
 Art tests, 236-239
 Ashbaugh, E. J., 163*n*.
 Attenborough, J., 272
 Audiometer, Maico D-6, 183*n*.
 Ayres, L. P., 148*n*.
 Ayres Measuring Scale for Handwriting, Gettysburg Edition, 148, 151-152
 Ayres Spelling Scale, 162-164

B

- Baker, F., 197*n*.
 Baker, H. J., 289*n*.

- Barnard, H. C., 3*n.*, 4*n*.
 Barr, A. S., 70, 318
 Bear, R. M., 189*n*.
 Betts, E. A., 186*n.*, 202
 Betts-Keystone telebinocular, 186
 Betts Ready-to-Read tests, 186-188
 Binet, A., 255
 Binet-Simon tests, 255-256
 Binocular reading graph, 190
 Bixler, H. H., 169*n*.
 Blin, 255
 Boas, F., 256
 Bohnhoff, L., 94
 Bolton, T. L., 256
 Bovard, J. F., 246
 Bowers, E. V., 180*n*.
 Bramkamp, J. D., 62*n*.
 Bridges, J. W., 257
 Briggs, T. H., 318
 Broer, M. R., 246
 Bronner, A. F., 277*n*.
 Broom, B. H., 92*n.*, 276*n.*, 300
 Broom, M. E., 12, 27*n.*, 30, 32, 34, 37, 41, 47, 56, 61, 62*n.*, 70, 73, 74, 75*n.*, 78*n.*, 91*n.*, 93*n.*, 121*n.*, 161*n.*, 197*n.*, 271*n.*, 272, 300, 301
 Brown, A. W., 232*n*.
 Brown, W., 71
 Brown-Woody Civics Test, 232-233
 Brueckner, L. J., 158, 168*n.*, 172, 182, 203, 223, 234, 318
 Buckingham, B. R., 134, 163*n.*, 259, 300
 Buckingham Extension of Ayres Spelling Scale, 163-164
 Buehrmann, A., 200*n*.
 Burgess, R. W., 32, 61
 Buros, O. K., 203
 Bursch, J. F., 134
 Burt, C., 256
 Burton, V. N., 232*n*,

- Burton, W. H., 232n., 318
 Burton Civics Test, 232
 Buswell, G. T., 208n.
 Buswell-John Diagnostic Chart for
 Fundamental Processes in
 Arithmetic, 208-211

C

- California Tests of Mental Matur-
 ity, 281, 289-294, 297
 Carey, G. L., 236n.
 Carroll, J. S., 70n.
 Cattell, J. M., 256
 Chaddock, R. E., 56, 58
 Chadwick, E. B., 4n.
 Characteristics of tests, 16-19
 (See also Objective tests, char-
 acteristics of; Standardized
 tests, characteristics of)
 Childs, H. G., 257
 Civics tests (see Social science tests)
 Clark, H., 93
 Clark, J. D., 98n.
 Clark, J. R., 221n.
 Clark, W. W., 168n., 185n., 197n.,
 201n., 219n., 289n., 292n., 293n.
 Classroom Tests, 134
 Coaching, 19
 Cole-Vincent Group Intelligence
 Test for School Entrants, 281
 College of Ephibi, 3
 Compass Diagnostic Tests in Arith-
 metic, 211-217
 Completion tests, 112-117
 Composition, mechanics, and gram-
 mar, teaching and testing of,
 172-174, 177, 182
 tests in, 172-182
 Composition scales, 173-176
 uses of, 176-177
 Conard, E. U., 150n.
 Conard Manuscript Writing Stand-
 ards, 149-151
 Conkling, F. R., 180n.
 Connor, W. L., 169n.

- Constructing Tests and Grading in
 Elementary and High School
 Subjects*, 107n., 108n., 126n.,
 134, 135
 Cooke, D. H., 37, 41
 Correlation, coefficient of, 38-42,
 68-71

- computation form for, 321
 Cozens, F. W., 246
 Cross, E. A., 160n.
Crossroads in the Mind of Man, 273
 Cunningham, B. V., 284n., 285n.

D

- Daly, P. F., 71
 Damaye, 255
 Dearborn, W. F., 273, 300
 DeGraff, M. H., 230n.
 Denny, E. C., 231n.
 Denny-Nelson American History
 Test, 231-232
 DeSanctis, S., 255
 DeSilva, F. W., 300
 Detroit Beginning First Grade In-
 telligence Test, 281
 Detroit Intelligence Tests, 289
 DeVault, N. M., 194n., 196n.
 DeVault Primary Reading Test,
 194-195
 DeVoss, J. C., 158, 172, 182, 223
 Diagnostic charts, in arithmetic, 210
 in handwriting, 157
 in oral reading, 193
 profile, 315
 in several subjects, 316
 in silent reading, 312
Diagnostic and Remedial Teaching,
 158, 172, 182, 203, 223, 234, 318
*Diagnostic Testing and Remedial
 Teaching*, 172, 182, 234
 Dickson, V. E., 298n., 300
 Difficulty, 18
 Discrimination, 19, 99-100
 Douglas, J., 91n.
 Drake, R. M., 241n.
 Drake Musical Memory Test, 241
 Dunn, A., 181n.

- Durrell, D. D., 201n.
 Durrell Analysis of Reading Difficulty Tests, 201
 Dvorak, A., 244n.
 Dvorak-Van Wagenen Analytical Scales of Attainment in Elementary Science, 244-245

E

- Ebbinghaus, H., 254
 Education, definition of, 1
 meaning of, 1-2
 Educational age, 20
Educational Diagnosis and the Measurement of School Achievement, 12
Educational Measurement, Fundamentals of, 12
Educational Measurements, 134, 273, 300
Educational Measurements and the Classroom Teacher, 12
Educational Measurements in the High School, 11, 92n., 101n.
 Educational measurements, in classroom teaching, 7-10, 105, 314-317
 in guidance, 7-10
 in supervision, 10-11, 310-314
 uses of, in administration, 7, 308-310
Educational Psychology, An Introduction to, 273
 Educational quotient, 20
Educational Statistics, 30, 37, 47
Educational Statistics for Beginning Students, 27n., 30, 32, 34, 37, 41, 47, 56, 58, 61, 70, 74
Educational Tests and Measurements, 158, 172, 182, 223
Educational Tests, The Use and Interpretation of, 32
 Emig, D., 79n., 115n., 126n.
 Epler, N., 159n.
 Error analysis, chart, 312
 procedure for, 72-73, 109

- Errors, constant, 21, 101
 variable or chance, 21, 101
 Essay tests, 108
 Experimental coefficient, 42-44, 71-72
 computation form for, 322
Experimental Psychology, The Foundations of, 274

F

- Farber, M., 272
 Farrand, L., 256
 Fenton, N., 134, 273, 300
 Fisher, Rev. George, 4
 Fletcher, J. M., 273
 Foods tests, 93-95
 Foster, J. C., 257
 Franseen, C. E., 178n.
 Franseen Diagnostic Tests in Language, 178
 Freeman, F. N., 154n., 273, 277n., 298n., 300
 Freeman Chart for Diagnosing Faults in Handwriting, 154-155
French Tradition in Education, The, 3n.
 Frequency polygon, 44, 45
 Furfey, P. F., 71

G

- Galton, F., 254, 256
 Garrett, H. E., 28n., 30, 37, 41, 57, 71, 72, 272, 273, 277n.
 Gates, A. I., 195n., 199n., 201n., 203, 245n.
 Gates Reading Diagnosis Tests, 201-202
 Gates Silent Reading Tests, Primary Series, 195-196
 Gates Series for Grades 3 through 8, 199-200
 Gates-Strang Health Knowledge Test, 245
 General science tests, 69-70, 98-101
 validity of, 69-70

- Geography tests (*see* Social science tests)
- German Education*, 4n.
- Gestalt psychology, 260, 264-265
- Gibbon, B. H., 91n.
- Gilbert, J. F., 256
- Gilliland, A. R., 12
- Glassow, R. B., 246
- Goddard, H. H., 256
- Grade placement score, 20
- Graham, F. B., 169n.
- Grammar (*see* Composition, mechanics, and grammar)
- Graphs, 44-47
- Gray, W. S., 191n.
- Gray Standardized Oral Reading Check Tests, 191-194
- Greene, H. A., 32, 167n., 211n., 230n., 246
- Greenwich Hospital School, 4
- Gregory, C. A., 12, 178n., 229n.
- Gregory, H., 178n.
- Gregory, M. A., 178n.
- Gregory-Spencer Geography Tests, 229-230
- Griffith, C. R., 273
- Griffiths, N. L., 186n.
- Grosvenor, E. L., 272
- Growth of the Mind, The*, 265
- Guidance, measurement and, 7-10
- Guidance, Principles of*, 318
- Guiler, W. S., 159n.
- Guy, J. F., 168n.
- H
- Hagerty, M. E., 229n.
- Handling Test Scores, Methods of*, 74
- Handwriting, skills necessary in, 147-148
- Handwriting tests and scales, 147-158
- diagnostic materials for measuring quality of handwriting, 154-158
- measuring handwriting quality, 148-154
- measuring manuscript handwriting quality, 148-151
- Handwriting tests and scales, measuring speed of handwriting, 148
- validity of, 69
- Hardaway, M., 63n.
- Hardwick, R. S., 257
- Harrison, M. L., 203, 299n.
- Harvill, A., 129n.
- Haskins, C. H., 3n.
- Hause, E., 301
- Health tests, 245
- Healy, W., 277n.
- Heredity and Environment*, 274
- Herrick, C. J., 265n.
- Herring, J. P., 257
- Hight, M. O., 97n., 160n.
- Hildreth, G., 186n.
- Hill, M. F., 257
- Histogram, 44, 45
- Historical Introduction to Modern Psychology*, 273
- History of Modern Education, The*, 4n.
- History tests (*see* Social science tests)
- Hoke, K. J., 92n., 134, 158, 172, 182, 203, 223, 234, 246, 274, 300, 318
- Holzinger, K. J., 30, 34, 37, 41, 57, 61, 71
- How to Experiment in Education*, 42
- How to Measure*, 92n., 134, 158, 172, 182, 203, 223, 234, 246, 274, 300, 318
- How to Measure in Education*, 12, 42, 318
- Hudelson, E., 174n.
- Hudelson English Composition Scale, 174-175
- Huey, E. B., 257
- Hull, C. L., 259, 260, 300
- Hunkins, R. V., 159n.
- Hunt, T., 272
- I
- Identification tests, 127-129
- Improvement of Reading, The*, 201n., 203

- Improving Instruction*, 318
 Imus, H. A., 189*n*.
Individual Mental Tests and Testing, A Manual of, 277*n*.
Industrial Education, Tests and Measurements in, 246
 Ingraham, J., 197*n*., 198*n*.
 Intelligence, abstract or academic, 268-270, 272
 aesthetic, 268-269
 area of, 267
 characteristics of, 266-268
 concrete, 268, 270-272
 definitions of, 258-261
 kinds of, 268-272
 level of, 266
 psychological theories of, 261-266
 range of, 266-267
 social, 268-269, 271-272
 speed, factor in, 267-268
Intelligence, Brain Mechanisms and, 261*n*., 273
Intelligence, Early Conceptions and Tests of, 253*n*., 274
 Intelligence quotient, 20
 constancy of, 21, 298
 Intelligence testing, philosophy of, 275-277
Intelligence Testing, 277*n*.
Intelligence Testing—Methods and Results, 264*n*., 274, 278*n*., 298*n*., 300
Intelligence Tests, 273, 300
 Intelligence tests, construction of, 278-280
 history of, 253-258
 intermediate grade, 285-294
 primary grade, 283-285
 scores yielded by, 296-298
 specific types of content of, 280-283
 types of, 277-278
 upper grade, 285-294
 uses of, 299
 validity of, 78, 294-296
 Iowa General Information Test in American History, 230-231
 Iowa Spelling Scales, 163-166
 Iowa Spelling Test, 167-168
 Issacs, R., 272
- J
- Jastrow, J., 256
 Jerome, H., 61
 Jesuits, 4
 John, L., 208*n*.
 Jones, A. J., 318
 Jordan, A. M., 276*n*., 300
 Jordan, R. H., 12
 Jorgensen, A. N., 32
 Jorgenson, C., 273
 Judd, C. H., 269
- K
- Kelley, T. L., 30, 42, 160*n*., 171*n*., 234, 261, 273, 300
 Kellogg, W. N., 272
 Kelly, F. J., 158, 172, 182, 223
 King, F., 93
 Kinter, M., 246
 Kline, L. W., 236*n*.
 Kline-Carey Measuring Scale for Freehand Drawing, 236-237
 Knight, F. B., 211*n*.
 Knuth W. F., 242*n*.
 Knuth Achievement Tests in Music, 242-243
 Koffka, K., 264, 265*n*.
 Köhler, W., 260
 Kohs, S. C., 257
 Kraepelin, E., 254
 Kramer, L., 94
 Krey, A. C., 234, 300
 Kuhlmann, F., 257, 286*n*., 287*n*., 300
 Kuhlmann-Anderson Intelligence Tests, 280, 286-288
 Kwalwasser, J., 241*n*., 243*n*., 246
 Kwalwasser-Ruch Test of Musical Accomplishment, 241-242
 Kwalwasser Test of Musical Information and Appreciation, 243-244
- L
- Lane, J. B., 159*n*.
 Language, uses of, 146

- Language arts, tests in, 172-203
 Lanier, L. H., 273
 Lashley, K. S., 261, 265, 273
 Lauderbach, J. C., 301
 Leamer, E. W., 152*n*.
 Leamer Scale for Graded Standards in Quality of Handwriting, 152-154
 Lee, J. M., 5*n*., 11, 185*n*., 318
 Lee-Clark Reading Readiness Test, 185
 Lehman, H. C., 74
 Leonard, J. P., 180*n*.
 Lewerenz, A. S., 237*n*.
 Lewerenz Tests in Fundamental Abilities of Visual Art, 237-238
 Limitations of tests (*see* Objective tests, limitations of; Standardized tests, limitations of)
 Lincoln, E. A., 72
 Los Angeles Elementary Reading Test, 197-198
 Los Angeles Diagnostic Tests: Fundamentals in Arithmetic, 219
 Reasoning in Arithmetic, 218-219
 Los Angeles Primary Word Recognition Test, 194
 Lovelace, L. H., 121*n*.
 Lowe, G. M., 277*n*.
- M
- McAnulty, E., 197*n*., 201*n*.
 McBroom, M., 203
 McCall, W. A., 12, 44, 170*n*., 285*n*., 286*n*., 318
 McCall Multimental Scale, 285-286
 McCarthy, C., 128*n*., 296*n*.
 McCloy, C. H., 246
 McFadden, J. H., 259
 McGurn, A., 78*n*.
 MacLatchy, V. R., 273
 MacQuarrie, T. W., 273
 Mansbridge, A., 4*n*.
 Manuscript writing standards, 149-151
 Matching tests, 126-127
Mathematics of Statistics, An Introduction to the, 32, 61
Matrix of the Mind, The, 258*n*.
Maze Test and Mental Differences, The, 278*n*.
 Measurement, in administration, 308-310
 and guidance, 7
 purposes of, 2-3, 5-6
 in recent education, 4-5
 in schools of past, 3-4
 and supervision, 10-11, 301-314
 and teaching, 6-10, 314-317
Measurement of Intelligence, The, 264*n*., 274
Measurement in Secondary Schools, A Guide to, 5*n*., 11, 318
Measuring Intelligence, 278*n*.
Measuring Results in Education, 11, 12, 172, 300
 Mechanics of English (*see* Composition, mechanics, and grammar)
 Median, 32-34, 57-58
 computation form for, 320
 Meier, N. C., 238*n*.
 Meier-Seashore Art Judgment Test, 238-239
 Melby, E. O., 158, 172, 182, 203, 223, 234, 318
 Meltzer, H., 134
 Mental age, 20
Mental Measurement, The Essentials of, 71
 Mental tests (*see* Intelligence tests)
Mental Tests, 273, 277*n*., 298*n*., 300
Mental Tests, A Handbook of, 300
Mental Tests and the Classroom Teacher, 298*n*., 300
Mental Tests in Clinical Practice, 300
 Merrill, M. A., 257, 278*n*.
 Metropolitan Achievement Grammar Tests, 179
 Metropolitan Reading Readiness Test, 186
 Meumann, E., 254
 Miller, G. F., 135
Minimum Essentials of Statistics, 37,

- Mitchell, B. C., 30
 Monroe, W. S., 11, 71, 158, 172,
 182, 223
 Morrison, J. C., 170n.
 Morton, R. L., 159n.
 Mudgett, B. D., 47
 Multiple-response tests, 122-126
 Murchison, C., 274
 Murphy, E. J., 181n.
 Murphy, G., 273
 Murphy, L. C., 200n.
Music, Tests and Measurements in,
 246
 Music teaching, purpose of, 239
 Music tests, 74, 239-244
 validity of Seashore test, 74

N

- Nature of "Intelligence" and the
 Principles of Cognition, The,*
 263n.
 Nelson, M. J., 231n.
*Neurological Foundations of Animal
 Behavior,* 265n.
 Newkirk, L. V., 246
*Newer Practices in Reading in the
 Elementary School,* 203
 Nielsen, N. P., 246
*1938 Mental Measurements Year-
 book, The,* 203
 Norm, defined, 14n.
 limitations of, 102
 uses of, 102
 Nowlin, G., 123n.
 Nygaard, P. H., 71

O

- Objective or New-type Examination,
 The,* 11, 87n., 108n., 112n.,
 117n., 134, 135
Objective Tests, 134
 Objective tests, advantages and
 limitations of, 102-103, 105-
 112
 characteristics of, 105-107
 preparation of, 131-133

- Objective tests, simple classification
 of types of, 107
 (See also Completion tests;
 Identification tests; Matching
 tests; Multiple-response tests;
 Rearrangement tests; Simple
 recall tests; True-false tests)
 Objectivity, 18, 87-88, 106
 Odell, C. W., 11, 32, 35, 37, 42, 44,
 47, 57, 71, 92n., 101n., 134
 Oehrn, A., 254
 Ogden, R. M., 260, 264
 Ogive, 46
Older Universities of England, The,
 4n.
 Ophthalmograph, 188-191
 Oratorian Schools, 4
 Orleans, J. S., 134, 135, 159n., 170n.
 Otis, A. S., 288n., 289n.
 Otis Quick-scoring Mental Ability
 Tests, 281, 282, 288-289
 Oxford University, 4

P

- Paterson, D. G., 135, 257
 Paulsen, F., 4n.
 Paulu, E. M., 172, 182, 234
 Pearce, B., 180n.
 Pearson, C., 256
 Percentile curve, 46, 65
 Percentiles, computation of, 33
 use of, 63-65
 Performance, pupil, characteristics
 of, 21
 defined, 21
 Perry, F. V., 94n.
 Perry, R. C., 300
 Peters, C. C., 58
 Peterson, J., 253n., 273, 274
 Phillips, D. P., 159n.
*Physical Education, Measuring
 Achievement in,* 246
*Physical Education, Tests and Meas-
 urements in,* 246
*Physical Education Activities,
 Achievement Scales in,* 246

- Pintner, R., 264, 273, 274, 278*n.*, 283*n.*, 284*n.*, 285*n.*, 298*n.*, 300
- Pintner-Cunningham Primary Mental Test, 280, 284-285
- Pintner Non-language Primary Mental Test, 283-284
- Porter, M. P., 301
- Porteus, S. D., 257, 258, 260, 278*n.*
- Potter, M. A., 221*n.*
- Pressey, L. C., 74, 155, 180*n.*, 182, 223, 300
- Pressey, S. L., 74, 155, 179*n.*, 180*n.*, 182, 223, 300
- Pressey Chart for Diagnosis of Illegibilities in Handwriting, 155-156
- Prewit, I., 148*n.*
- Profile chart, 315, 316
- Progressive Achievement Spelling Test, 168-169
- Psychological Tests, Methods, and Results*, 277*n.*
- Psychology, Dictionary of*, 21
- Psychology and Education*, 260*n.*
- Psychology in Education*, 273
- Psychology, Great Experiments in*, 273
- Psychology of Elementary School Subjects*, 147*n.*, 158, 172, 223
- Psychology of Secondary Education*, 269*n.*
- Public School Achievement Spelling Test, 170-171
- Publishers of tests, 319
- Reading Difficulties, The Prevention and Correction of*, 202
- Reading Readiness*, 203, 299*n.*
- Reading, teaching and testing of, 183-185
- tests, 88-91, 182-203
- validity of, 79, 89-91, 197
- Reading readiness tests, 185-188
- Rearrangement tests, 129-131
- Reed, H. B., 147*n.*, 158, 172, 223
- Reidy, A., 221*n.*
- Reliability, 17-18, 87-88
- of algebra tests, 92-93
- of civics tests, 81
- of foods tests, 94-95
- of general science tests, 99-101
- of mental tests, 284-286, 288-289, 292-293, 295-296, 298
- of reading tests, 90-91
- of spelling tests, 96-98
- Research for Teachers*, 134, 259*n.*, 300
- Rice, G. D., 94*n.*, 107*n.*, 135
- Rice, J. M., 5
- Rinsland, H. D., 107*n.*, 108*n.*, 126*n.*, 134, 135
- Rothney, J. W., 189*n.*
- Ruch, G. M., 11, 87, 92*n.*, 94*n.*, 100*n.*, 107*n.*, 108*n.*, 112*n.*, 117*n.*, 134, 135, 160*n.*, 171*n.*, 211*n.*, 214*n.*, 230*n.*, 234, 241*n.*
- Rudd, M., 91*n.*
- Ruhlen, H., 180*n.*
- Russell, C., 11, 12, 134

Q

Quartile deviation, 37, 61-62

R

Randall, F. B., 301

Range, 58-59

Ratio, 20

Raubenheimer, A. S., 273

Reading, Controlled, 203

Reading, The Teaching of: A Second Report, 203

S

Sangren, P. V., 12, 198, 221*n.*, 294*n.*, 318

Sangren-Reidy Survey Tests in Arithmetic, 221-222

Sangren-Woody Reading Test, 198-199

Scale, 14

Schneck, M. R., 277*n.*

School, as educative agency, 2

Schorling, R., 221*n.*

Schorling-Clark-Potter Arithmetic Test, 221

- Schwesinger, C., 274
- Science, in elementary school, 244
- Science tests, 244-245
- Scores, derived, 18-21
- Scoring, of tests, 107, 114-116, 121, 124
- Scoring keys, 114, 116, 287
- Scudder, C. R., 273
- Sealy, G. A., 134
- Seashore, C. E., 238*n.*, 240*n.*
- Seashore Tests of Musical Talent, 74, 240
- Seaton, J. T., 179*n.*
- Seaton-Pressey Diagnostic Tests in English Composition, 179-180
- Shimberg, M. E., 277*n.*
- Shotola, I., 94
- Shotwell, A. M., 11, 134, 158, 172, 182, 203, 223, 234, 246, 300
- Simon, T., 255
- Simple recall tests, 112-117
- Sims, B., 115*n.*
- Skiff, J. M., 82*n.*, 276*n.*
- Smith, H. L., 182, 203, 233, 234, 300
- Smith, L. A., 81*n.*
- Social science, teaching and testing of, 225-226, 233-234
- Social science tests, 225-234
- civics tests, 232-233
- validity of, 81
- geography tests, 226-230
- history tests, 230-232
- Social Sciences, Tests and Measurements in the*, 234, 300
- Sorenson, H., 32, 35, 37, 42
- Spearman, C., 256, 259, 261, 262, 263*n.*, 264, 274
- Specimen Objective Examinations*, 94*n.*, 107*n.*, 135
- Spelling tests, 5, 95-98, 158-172
- experiments with, 158-162
- validity of, 96-98, 161-162
- Spencer, P. I., 220*n.*, 229*n.*
- Spencer Diagnostic Arithmetic Tests, 220-221
- Stagner, R., 273
- Standard deviation, 35-37, 59-61
- computation form for, 320
- Standard Tests*, 11, 12, 134
- Standard Tests in the Elementary School*, 11, 134, 158, 172, 182, 203, 223, 234, 246, 300
- Standardized tests, advantages and limitations of, 101-102
- (See also Composition scales; Grammar tests; Handwriting scales and tests; Reading tests; Spelling tests)
- Statistical data, classification of, 28-29
- collection of, 28
- tabulation of, 28, 29
- Statistical Method* (Jerome), 61
- Statistical Method* (Kelley), 30, 42
- Statistical Method in Education*, 32, 35, 37, 42, 44, 47, 57, 71
- Statistical Methods for Students in Education*, 34, 37, 41, 57, 61, 71
- Statistical Procedures and Their Mathematical Bases*, 58
- Statistical Tables and Graphs*, 47
- Statistics, defined, 26
- needed by teacher, 27-28
- Statistics, Exercise Manual in*, 30
- Statistics, The Fundamentals of*, 35, 47
- Statistics, Principles and Methods of*, 56, 58
- Statistics in Psychology and Education*, 28*n.*, 30, 37, 41, 57, 71, 72
- Statistics for Students of Psychology and Education*, 32, 35, 37, 42
- Steeves, H. R., 160*n.*
- Stein, M. L., 273
- Stenquist, J. L., 273
- Stern, W., 254, 258, 260
- Stoddard, G. D., 92*n.*, 100*n.*, 234
- Stone, C. R., 200*n.*
- Stone, C. W., 5
- Stone Narrative Reading Tests, 200, 311-312
- Strang, R., 245*n.*, 273
- Streator, N., 93
- Strong, A. C., 257
- Studebaker, J. W., 211*n.*
- Stuit, D. B., 71

344 MEASUREMENTS IN THE ELEMENTARY SCHOOL

Subject age, 20
 Sullivan, E. T., 289*n.*, 292*n.*, 293*n.*
 Summary graph, 46-47, 67
 Supervision, measurement and, 10-11, 310-314
Supervision of Instruction, The, 318

T

Taylor, E. A., 190, 203
 Taylor, J. F., 300
 Telebinocular, 186
 Tenwolde, H., 69*n.*
 Terman, L. M., 160*n.*, 171*n.*, 257, 259, 260, 278*n.*

Terman Group Test of Mental Ability, 282

Terminology, 14-21

Test, achievement, 14
 cycle, 15
 definition of, 14
 diagnostic, 14
 duplicate forms of, 14-15
 mental, 15
 omnibus, 14
 power, 15
 rate, 15
 survey, 14

(See also Intelligence tests; Objective tests; Standardized tests)

Testing Program, Administration of the, 12, 318

Tests, early standardization of, 5

Tests and Measurements, 182, 203, 223, 234, 300

Tests and Measurements for Teachers, 92*n.*, 134, 274

Tests and Measurement in High School Instruction, 92*n.*, 100*n.*, 234

Theory of Educational Measurements, An Introduction to the, 11, 71

Theory of Mental and Social Measurements, An Introduction to the, 5*n.*

Theory of Statistics, An Introduction to the, 30, 37, 42, 58

Thomson, G. H., 71, 256

Thorndike, E. L., 5, 263-266, 274

Thurstone, L. L., 35, 47, 261

Tiegs, E. W., 92*n.*, 134, 168*n.*, 274, 289*n.*, 292*n.*, 293*n.*

Torgerson, T. L., 170*n.*

Trabue, M. R., 11, 12, 172, 300

Trachsel, M., 78*n.*

Traditional examinations (see Essay tests)

Traditional Examinations and New-type Tests, 134

Trilling, M., 93

True-false tests, 117-122

U

Uhl, W. L., 12

Universities, The Rise of, 3*n.*

Universities of Ancient Greece, The, 3*n.*

University of Bologna, 3

University of Paris, 3

Upshall, C. C., 273

Upton, C. B., 215*n.*

Upton Inventory Test in Arithmetic, 215, 218

Use of Standardized Tests, Introduction to the, 182, 223, 300

Uses of tests (see Educational measurements, uses of)

V

Validity, 15-16, 87-88

 of algebra tests, 92-93

 of civics tests, 81

 of foods tests, 94-95

 of general science tests, 69-70, 99-100

 of handwriting scales, 69

 of mental tests, 78, 285, 286, 289, 293-296

 of reading tests, 79, 89-91, 197

 of spelling tests, 96-98

Van Voorhis, W. R., 58

Van Wagenen, M. J., 12, 168*n.*, 185*n.*, 244*n.*

- Van Wagenen Reading Readiness Test, 185-186
- Variable, defined, 28
- W
- Walden, J. W. H., 3*n*.
- Wallin, J. E. W., 257
- Walther, E. C., 226*n*.
- Warren, H. C., 21
- Webb, C. W., 75*n*.
- Webb, L. W., 11, 134, 158, 172, 182, 203, 223, 234, 246, 300
- Weller, L. E., 160, 161*n*., 197*n*.
- Wells, F. L., 300
- Wert, J. E., 30, 37, 47
- West, P. V., 156*n*., 157
- West Chart for Diagnosing Elements of Handwriting, 156-158
- Whipple, G. M., 203
- Wiedfeld, M. T., 226*n*.
- Wiedfeld-Walther Geography Test, 226-229
- Williams, A. J., 195*n*.
- Williams, S. G., 4*n*.
- Williams Primary Reading Test, 195-196
- Willing, M. H., 175*n*.
- Willing Scale for Measuring Written Composition, 175-176
- Wilson, G. M., 92*n*., 134, 158, 172, 181*n*., 182, 203, 223, 234, 246, 274, 300, 318
- Wilson Language Error Test, 181-182
- Winch, W. H., 257
- Witty, P. A., 74, 300
- Wood, B. D., 160*n*.
- Wood-Jones, F., 258*n*.
- Woody, C., 12, 198, 232*n*., 318
- Worcester, D. A., 134, 273, 300
- Wright, L. A., 300
- Wright, W. W., 182, 203, 223, 234, 300
- Wundt, W., 254
- Y
- Yerkes, R. M., 257
- Young, K., 274
- Yule, G. U., 30, 37, 42, 58