MEASUREMENT AND ADJUSTMENT SERIES

EDITED BY LEWIS M. TERMAN

EDUCATIONAL MEASUREMENT IN THE ELEMENTARY GRADES

BY I. N. MADSEN, PH.D.

Director of the Department of
Tests and Measurements, State Normal School
Lewiston, Idaho

WORLD BOOK COMPANY

Yonkers-on-Hudson, New York

WORLD BOOK COMPANY
THE HOUSE OF APPLIED KNOWLEDGE

Established 1905 by Caspar W. Hodgson

YONKERS-ON-HUDSON, NEW YORK
2126 PRAIRIE AVENUE, CHICAGO

Also BOSTON : ATLANTA : DALLAS
SAN FRANCISCO : PORTLAND : MANILA

"UNTO one he gave five talents, to another two, to another one; to every man according to his several ability." As old as Scripture is recognition of the fact that capacities vary, but new as our generation is the application of scientific method to the determination of individual differences. Much has been done in the way of providing tests for students in the elementary grades of our schools, and many tests for making various classifications are now available; but too often teachers do not understand the theory underlying the tests that they use; too often they do not know how to get the maximum benefit from testing. For the guidance of teachers who wish to be informed on the technique of testing in order that they may render the best possible service to their charges, this book on *Educational Measurement in the Elementary Grades* has been prepared.

MAS : MEMEG-2

PRINTED IN U.S.A.

PREFACE

THERE are now many excellent books available in the field of educational measurements. None, however, deals comprehensively and solely with all the major phases of testing for the elementary grades. In writing this book the author has had in mind the training of teachers, supervisors, principals, and others whose chief concern will be with these grades. It is here that we find a great majority of the school children, as well as the greatest development of standardized tests. Yet discouragingly little use is made of such tests for the improvement of teaching. A splendid tool lies idle or is unskillfully used because of the inadequacy of the training given craftsmen.

Many normal schools and teachers' colleges are attempting to remedy this situation by including in their curricula required courses in educational tests and measurements. But it is usually not possible to devote to these more than twelve weeks, with three or four hours of instruction a week. Most students in these institutions are certified to teach after two years, or less, of training. This means that instruction in educational measurements, as in other subjects, must be limited to the minimum essentials. An attempt has therefore been made to include in this book only topics which have been found most significant for the elementary grades.

During ten years of experience in teaching beginning students, the author has become convinced that the best results are obtained by pointing out, as definitely and concretely as possible, the relation of testing to such other important phases of teaching as the course of study, the determination of objectives and their attainment, pupil diagnosis, remedial instruction, sectioning classes, classifying pupils, etc. Wherever possible, this relationship has been pointed out. Fortunately an abundance of laboratory material is available to supplement the theoretical work in a measurements course. Thus the students can participate in giving

the tests, they can score the results, tabulate or otherwise treat results to provide a basis for their interpretation, etc.

A discussion of statistical concepts has been placed in the first part of the book, since a relatively early mastery of these concepts will greatly facilitate mastery of the later chapters. The instructor should find no difficulty in obtaining local material to demonstrate the usefulness of statistical concepts and to provide the necessary drill or practice in applying them. It is hoped that Chapters V and VI may give the student an appreciation of the importance of the relation between intelligence and progress in school and that they may also make possible more intelligent participation in testing programs conducted by experts. It may be said that many teachers misinterpret intelligence tests; however, the same can be said about achievement tests and, for that matter, about any other method of pupil investigation. Should valuable but complicated tools be discarded because of clumsy workmen? Is it not more logical to insist on better training and higher standards? In selecting achievement tests for discussion in Chapters VII and VIII, the aim was to get a cross-section of available tests of this type rather than to apply rigid criteria of merit. While there is no doubt that great progress has been made in the evaluating of tests, we have not yet reached the point where the doctors agree.

The author is indebted to many who have helped directly or indirectly in the preparation of this book. He is particularly indebted to Dr. Lewis M. Terman for many helpful suggestions. The inspiration that he has received from the work of all the contributors to the theory and practice of educational measurement is as gratefully acknowledged; wherever possible, specific acknowledgment has been made in the text to individuals of this group who have been directly or indirectly quoted.

I. N. MADSEN

EDITOR'S INTRODUCTION

THIS book by Dr. Madsen, *Educational Measurement in the Elementary Grades*, fills a long-existing gap in the *Measurement and Adjustment Series*. On a subject which is so new and which has so many technical aspects it is not easy to write a book suitable for the average teacher or for the beginning student in normal schools and teachers' colleges. In attempting to prepare such a book the writer is faced by two dangers. If he writes a book that is readily comprehensible, he is likely to omit or slight important technical considerations without an understanding of which the results of educational measurements cannot be properly interpreted. If on the other hand he is careful to avoid this danger, the result is likely to be a book better adapted to the graduate student than to the undergraduate or the teacher-in-service.

Of the half-dozen or more manuscripts which have been submitted to this series as strictly introductory texts in educational measurement, this by Dr. Madsen is the first that the Editor has been willing to recommend for publication. The author himself would be the last to claim that this book is free from faults, but it can fairly be said that it provides an excellent orientation to the student who is entering upon the subject for the first time. In the opinion of the Editor, Dr. Madsen's book gives about all the information on this subject that can reasonably be regarded as essential for the rank and file of teachers to have. It is rather generally agreed that all who are preparing to teach should have at least one general course dealing specifically with educational tests as distinct from intelligence tests. However, to attempt to make every teacher an expert in the use and interpretation of educational measurements would be a mistake; there are too many other things which it is important for the teacher

to get, not the least of which is a fair degree of expertness in the teaching process.

As is clearly indicated by its title, this book has been prepared exclusively for those who teach or expect to teach in the elementary grades. There is no need of burdening the elementary teacher with information about measurement methods designed for use in the high school or college grades. Those fields have been admirably dealt with by two earlier texts in this series; namely, *Tests and Measurements in High School Instruction*, by Ruch and Stoddard, and *Measurement in Higher Education*, by Wood. A splendid critical treatment of the whole subject of educational testing, for the more advanced student, will be found in Kelley's *Interpretation of Educational Measurements*, also in this series.

The content and organization of this book have gradually taken form over a period of several years in teaching the subject to students in a normal school. The book contains no material which has not been extensively tried out with students of the academic level for which it is designed. It is this fact, doubtless, which is responsible for the author's simple style and straightforward exposition.

LEWIS M. TERMAN

CONTENTS

EDUCATIONAL MEASUREMENT IN THE ELEMENTARY GRADES

CHAPTER ONE

The Measurement of Individual Differences

Origin and development of the testing movement. Individual differences in human beings have probably always been noted and remarked upon. However, it is only during recent times that scientific observations and measurements of such differences have been made, and it is even more recently that the measurement of these differences has been made of practical use in education. In 1879 Wundt established the first laboratory for the scientific study of psychology. A few years later, in 1884, Sir Francis Galton founded his laboratory for anthropometric measurements. Long before this, however, Galton had been interested in the study of the heredity of mental traits and capacities, and in 1869 had published his *Hereditary Genius*. The testing movement was definitely introduced into America as early as 1890 by an American psychologist, J. McKeen Cattell, who had conducted experiments in Wundt's laboratory and who had later been associated with Galton. At this time Cattell outlined a testing program which was published in the English journal, *Mind*, and during this period he also began experimenting with tests for the measurement of mental traits.

For the fifteen years following 1890 psychologists were busily experimenting with tests that would effectively describe human intelligence. It was not until 1905, however, that Alfred Binet published his first rough scale; this he standardized on an age basis in 1908, and revised in 1911.

Coincident with these experiments in the measurement of mental traits, we find a movement which concerned itself with the progress and achievement of pupils in school. One phase of this movement dealt with the age-grade status of pupils and their elimination from school. Among the best-known investigations of this kind are those by Thorndike,[1] Ayres,[2] and Strayer,[3] published 1907, 1909, and 1911 respectively. These studies not only aroused widespread interest and discussion but resulted in a flood of similar investigations. In general, the purpose of such investigation was to ascertain the facts concerning age-grade status and elimination, to use these facts in determining the efficiency of the schools, to study causes for the conditions found, and to fix responsibility. The most striking fact revealed by these studies was the wide range of individual differences in the progress of pupils through school.

The second phase of the movement dealt with the development and the use of standardized tests. Dr. J. M. Rice,[4] as a result of his investigation of the efficiency in teaching spelling, is usually given credit for being the first to use comparative tests for measuring the results of teaching. Rice's investigation began in 1894 and attracted much attention for several years thereafter. During the year 1908 C. W. Stone,[5] under the direction of Professor Thorndike, developed a

[1] E. L. Thorndike, *The Elimination of Pupils from School* (United States Bureau of Education Bulletin, No. 4). Government Printing Office, Washington; 1907.

[2] Leonard P. Ayres, *Laggards in Our Schools.* Charities Publication Committee, New York; 1909. (Now published by Survey Associates, Inc., 112 East Nineteenth Street, New York.)

[3] George D. Strayer, *Age and Grade Census of Schools and Colleges* (United States Bureau of Education Bulletin, No. 5). Government Printing Office, Washington; 1911.

[4] J. M. Rice, "The Futility of the Spelling Grind," in *The Forum*, Vol. XXIII, pages 163–172, 409–419; 1897.

[5] C. W. Stone, *Arithmetical Abilities and Some Factors Determining Them* (Contributions to Education, No. 19). Teachers College, Columbia University, New York; 1908.

standardized arithmetic test; and in the course of the following year Thorndike's [1] handwriting scale appeared. It is apparent, therefore, that scientific methods for measuring the results of teaching, like those for measuring intelligence, are of recent origin.

The scientific measurement of human traits may conveniently be summarized under five heads: (1) anthropometric measurements, which are concerned with the measurement of different parts of the human body, such as height, weight, head girth, etc.; (2) measurements of sensory acuity, which measure the keenness of the sense organs — visual, auditory, tactual, etc.; (3) measurements of reaction time and motor ability, such as the speed of reaction to visual, auditory, or other sensory stimuli; (4) measurements of complex mental processes, such as memory, imagery, etc.; (5) measurements of general mental ability, special aptitudes, achievement in school, etc. It is this last group that has been found most significant in teaching, and consequently it will be our primary concern in this book.

The nature of individual differences. All the traits that are listed above, when measured, show wide individual differences. The nature of these differences can be best illustrated by arranging the measurement of a given trait in the form of a frequency table or curve. The following tables and curves are based upon actual measurement of the traits specified. The first table shows the range and distribution of height in a group of 202 women students in normal school. The same facts are represented graphically in Figure 1.

From Table 1 it will be seen that the 202 women composing the group range in height from 57 to 69 inches, that the mean (average) is 63 inches, that the greatest proportion of measurements cluster about the mean, and that the frequencies

[1] E. L. Thorndike, "A Handwriting Scale," in *Teachers College Record*, Vol. II; March, 1910.

TABLE 1

FREQUENCY TABLE SHOWING THE RANGE AND DISTRIBUTION OF HEIGHT IN A GROUP OF WOMEN STUDENTS

HEIGHT IN INCHES	FREQUENCY
69	1
68	3
67	8
66	19
65	26
64	32
63	36
62	29
61	24
60	15
59	6
58	2
57	1
Total	202
Mean	63

decrease in about the same way on either side of the mean. Results corresponding very closely to these would be obtained in any similar group of women. Indeed, accurate measurement of any one trait in such a group would reveal the same characteristic tendencies. These tendencies may be generalized as follows. (1) The abilities are continuous. They range without a break from the lowest to the highest. In Table 1 there appears to be a break of one inch between each group of frequencies and the group above or below. However, this is not a true break, since any one group of measures, such as the fifteen cases given opposite the 60 interval, include all the measures that range from 60 up to, but not including, 61. Sometimes breaks may appear to exist when a small

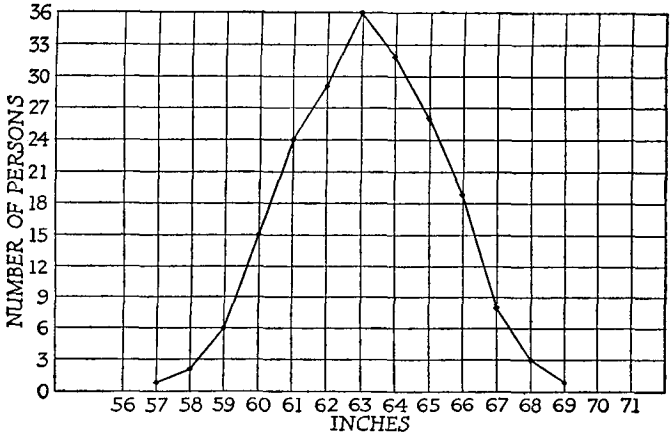

FIG. 1. Distribution of the heights of 202 women.

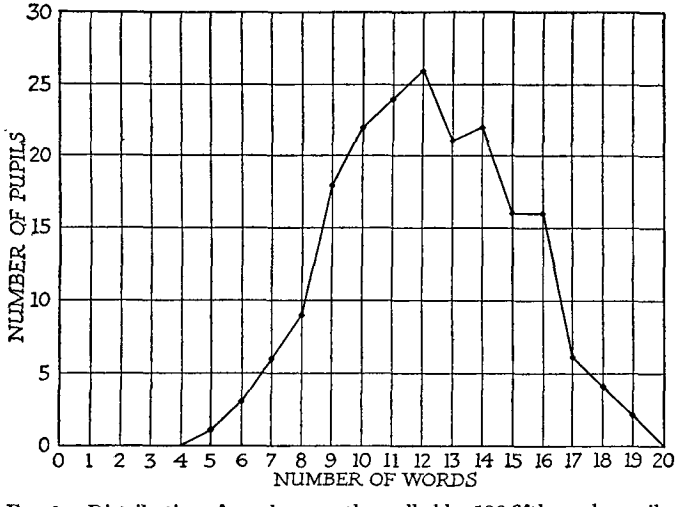

FIG. 2. Distribution of words correctly spelled by 186 fifth-grade pupils.

group is measured, but they are always filled in when additional measurements are made. (2) The abilities tend to cluster about the center of the distribution in such a way that a large proportion of measures are fairly close to the mean in magnitude. Thus, referring again to Table 1, we see that 97 women, or approximately one half of the total number, range in height from 62 to 64 inches inclusive. (3) The measures decrease in about the same proportion on both sides of the mean. When a frequency table is plotted in the form of a frequency curve, the result shows a tendency toward a symmetrical, bell-shaped curve. (4) Variations in a trait appear to be distributed in accordance with the law of chance as illustrated by coin tossing. This is significant in connection with the mathematical properties of the frequency curve

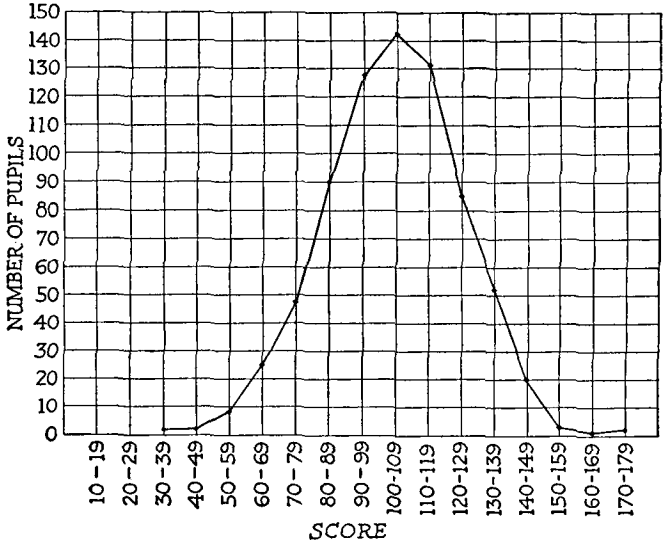

FIG. 3. Distribution of scores in Haggerty Intelligence Examination, Delta 2, by 737 seventh-grade pupils.

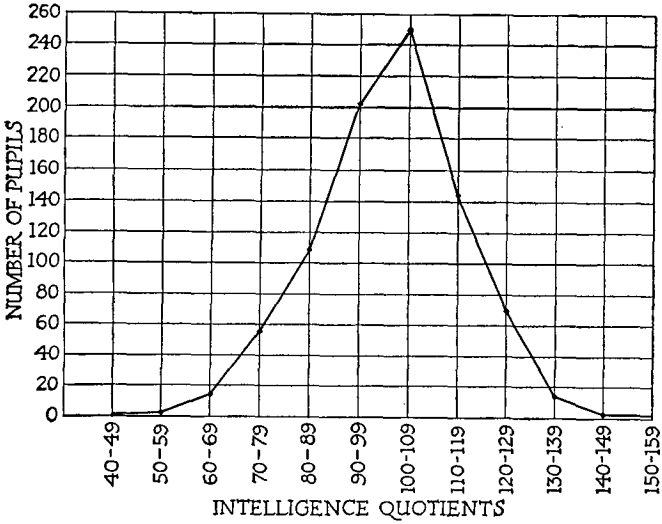

FIG. 4. Distribution of IQ's of 880 pupils according to the Stanford-Binet Test.[1]

and will be commented upon elsewhere. It should be emphasized that the characteristics listed above tend to appear only when the measurements are made accurately and with sufficient numbers. These tendencies are more concretely illustrated in the curves shown in Figures 1 to 4.

Importance of individual differences in school progress. It is obvious that individual differences, such as those represented in Figures 2 to 4, are of great importance in determining the progress of pupils through school. Largely because of these differences we find that pupils vary widely in their rate of progress and in the extent to which they profit from instruction. Starch quotes statistics from the St. Louis schools that are of interest in this connection. For many years these schools had promoted pupils at the end of each

[1] Adapted from I. N. Madsen, "Some Results with the Stanford Revision of the Binet-Simon Tests," in *School and Society*, May 10, 1924.