GRAW-HILL SERIES IN EDUCATION

MEASURING EDUCATIONAL ACHIEVEMENT

by **WILLIAM J. MICHEELS, Ph.D.**
Industrial Education Department
University of Minnesota

& M. RAY KARNES, Ph.D.
Industrial Education Department
University of Illinois

First Edition

New York Toronto London
McGRAW-HILL BOOK COMPANY, INC.
1950

Measuring Educational Achievement

McGRAW-HILL SERIES IN EDUCATION

HAROLD BENJAMIN, *Consulting Editor*

MEASURING EDUCATIONAL ACHIEVEMENT

Selected Titles from

McGRAW-HILL SERIES IN EDUCATION

HAROLD BENJAMIN, *Consulting Editor*

Allen · THE FEDERAL GOVERNMENT AND EDUCATION
Beaumont and Macomber · PSYCHOLOGICAL FACTORS IN EDUCATION
Bent and Kronenberg · PRINCIPLES OF SECONDARY EDUCATION
Bogue · THE COMMUNITY COLLEGE
Broom, Duncan, Emig, and Steuber · EFFECTIVE READING INSTRUCTION
Brubacher · MODERN PHILOSOPHIES OF EDUCATION
Brubacher · A HISTORY OF THE PROBLEMS OF EDUCATION
Butler and Wren · THE TEACHING OF SECONDARY MATHEMATICS
Butts · A CULTURAL HISTORY OF EDUCATION
Cook and Cook · A SOCIOLOGICAL APPROACH TO EDUCATION
Crow and Crow · MENTAL HYGIENE IN SCHOOL AND HOME LIFE
Croxton · SCIENCE IN THE ELEMENTARY SCHOOL
Davis · EDUCATIONAL PSYCHOLOGY
Davis and Norris · GUIDANCE HANDBOOK FOR TEACHERS
De Young · INTRODUCTION TO AMERICAN PUBLIC EDUCATION
Fedder · GUIDING HOMEROOM AND CLUB ACTIVITIES
Fernald · REMEDIAL TECHNIQUES IN BASIC SCHOOL SUBJECTS
Forest · EARLY YEARS AT SCHOOL
Good · DICTIONARY OF EDUCATION
Hammonds · TEACHING AGRICULTURE
Heck · THE EDUCATION OF EXCEPTIONAL CHILDREN
Hoppock · GROUP GUIDANCE
Kaulfers · MODERN LANGUAGES FOR MODERN SCHOOLS
Kaulfers, Kefauver, and Roberts · FOREIGN LANGUAGES AND CULTURES IN AMERICAN EDUCATION
Mays · PRINCIPLES AND PRACTICES OF VOCATIONAL EDUCATION
McCullough, Strang, and Traxler · PROBLEMS IN THE IMPROVEMENT OF READING
McKown · ACTIVITIES IN THE ELEMENTARY SCHOOL
McKown · HOME ROOM GUIDANCE
McKown and Roberts · AUDIO-VISUAL AIDS TO INSTRUCTION
Micheels and Karnes · MEASURING EDUCATIONAL ACHIEVEMENT
Mort · PRINCIPLES OF SCHOOL ADMINISTRATION
Mort and Reusser · PUBLIC SCHOOL FINANCE
Mort and Vincent · MODERN EDUCATIONAL PRACTICE
Mursell · DEVELOPMENTAL TEACHING
Mursell · SUCCESSFUL TEACHING
Myers · PRINCIPLES AND TECHNIQUES OF VOCATIONAL GUIDANCE
Schorling · STUDENT TEACHING
Schorling and Wingo · ELEMENTARY-SCHOOL STUDENT TEACHING
Sears · THE NATURE OF THE ADMINISTRATIVE PROCESS
Smith, Standley, and Hughes · JUNIOR HIGH SCHOOL EDUCATION
Sorenson · PSYCHOLOGY IN EDUCATION
Thorpe · PSYCHOLOGICAL FOUNDATIONS OF PERSONALITY
Thut and Gerberich · FOUNDATIONS OF METHOD FOR SECONDARY SCHOOLS
Warters · HIGH-SCHOOL PERSONNEL WORK TODAY
Wilson, Stone, and Dalrymple · TEACHING THE NEW ARITHMETIC
Winslow · THE INTEGRATED SCHOOL ART PROGRAM

PREFACE

This is a book for teachers. We hope it can serve as a handbook to be used and reused in answering the numerous questions on testing and evaluation that besiege the classroom and shop teacher.

The basic ideas and many of the suggested procedures have grown out of our combined experience in preparing testing materials for classroom use. We have been guided by our additional experiences in presenting such materials to teacher-training groups.

Throughout the book we have tried to keep uppermost in mind the busy teacher and especially the inexperienced instructor. In certain instances this has meant sacrificing the completeness and preciseness that would be necessary in a book written for the specialist in measurement. For example, standardized tests have not been accorded the treatment that is afforded them in many books on educational measurement. This is not to disparage the use of standardized tests. On the contrary, we feel they are vital factors in any effective program of evaluation. However, real understanding of the nature of such instruments is gained only after handling and using them. The typical teacher uses few, if any, standardized tests. Our efforts have been aimed at gathering suggestions that can be used by the teacher in constructing classroom tests and other measuring instruments. The basic principles discussed herein should help him to use wisely any standardized tests that he may administer.

As we see it, the major value of a book of this kind is to present the "how" of making and using tests and other instruments of appraisal. We have tried not to neglect the "what" and "why"; in fact, the introductory chapters are devoted to such discussion, although largely in overview fashion. However, the history of the testing movement, like the history of Ethiopia, becomes much more meaningful after you have been there.

The focal point in many instances has been industrial education, primarily because we know that area of instruction best. To preclude the inhibiting aspects of subject-matter specialization we

have added a rather large variety of examples from other subject-matter fields. Of course, the basic principles and procedures should be applicable in any teaching-learning situation.

We sincerely hope that this book will prove concretely useful and practically helpful both to the beginning and the experienced teacher who is desirous of making his evaluations more effective and meaningful.

We wish to express our sincere appreciation and thanks to the many students who have reacted to much of this material in draft form and have contributed a large share of the sample items; to John A. Butler, Walter W. Cook, Cyril Hoyt, John A. Jarvis, Verne L. Pickens, and H. T. Widdowson, who have read parts of the manuscript and offered valuable criticisms and suggestions; to Dean Horace T. Morse and Assistant Dean Alfred Vaughn of the General College, University of Minnesota, for permission to use selected items from the Comprehensive Examinations prepared for use in that college; to our colleagues in the Teacher-training Department, Armored School, Fort Knox, Kentucky, who provided much assistance in preparing certain preliminary materials that have been expanded or revised for inclusion herein; to the Bruce Publishing Company, publishers, and John J. Metz, editor, *Industrial Arts and Vocational Education,* for permission to use or adapt various articles prepared for that magazine by the authors and others; to David T. Ryans, executive secretary, and the Committee on National Teacher Examinations of the American Council on Education, for permission to adapt certain test items prepared by one of the authors for use in those examinations; to the other organizations and individuals, mentioned in footnote references, who allowed the use of cited materials; to Dr. Homer J. Smith and Professor Arthur B. Mays for advice and counsel as this project took shape and developed; and to Mrs. George E. DeVries and Mrs. J. B. Walden for their helpful assistance in the preparation of the manuscript.

WILLIAM J. MICHEELS
M. RAY KARNES

MINNEAPOLIS, MINN.
URBANA, ILL.
July, 1950

CONTENTS

CHAPTER ONE: AN INTRODUCTION TO MEASUREMENT

Take a good look at the following line. How long is it? Just for fun write your guess on a piece of paper. Do not use a ruler.

Do you think you came within one-eighth inch of the correct dimension? You can check by turning to the next page.

THE foregoing may seem like an odd way to start a discussion of achievement testing. It is. But it illustrates some important concepts about measurement—some basic ideas that should be understood from the beginning.

What Measurement Is

When you wrote down a figure, you illustrated to yourself what measurement is. You determined the length of the line in terms of a specified unit of measurement (inch). Measurement of any kind is a matter of determining how much or how little, how great or how small, how much more than or how much less than.[1] This is

[1] In more refined terms, to measure means to "observe or determine the magnitude of a variate" (*Encyclopedia of Educational Research,* p. 713). It should be noted that enumeration (counting) and measurement are not one and the same thing. In simplest terms, enumeration answers the question "how many." Measurement answers the question "how much."

true whether you are trying to measure the weight of a sack of potatoes, the color of a person's eyes, the radioactivity of an "atom-bombed" island, or the achievement of a student in the classroom.

Potatoes are usually measured in terms of *so many* pounds. A person's eyes contain *so much* blue or hazel or other color. An atom-bombed island is measured to determine *how much* radiation exists.

Achievement in the classroom is often stated in terms of *so many* points on a test or similar device that is supposed to measure the growth or achievement that has taken place.

Some of these things can be measured very accurately, others only roughly; still other things cannot presently be measured at all. This leads us to a basic precept of measurement and a consideration of what can be measured.

What Can Be Measured

Anything that exists at all exists in some quantity, and anything that exists in some quantity is capable of being measured.[2]

In a sense this statement is an expression of faith. It does not say that we are *now* able to measure all things. It does say that anything existing in quantity is capable of being measured. As a beginning test maker, it is very necessary that you have an understanding of this concept. You will then realize that certain achievement may be impossible of fine measurement at the present time, but you will also understand that the problem is one of devising or improving measuring instruments so that the job can be done in a better manner. In terms of teaching effort it is directly related to the problem of developing tests and other devices that will do a better job of measuring the outcomes of learning.

Fifty years ago there was no such thing as a Geiger counter for measuring the radioactivity of an area. Seventy-five years ago we had no standardized tests for measuring intelligence. One hundred years ago who would have thought of measuring length in terms of

---

The length of the line on page one is $3\frac{5}{8}$ inches.

---

[2] E. L. Thorndike, "The Nature, Purposes and General Methods of Educational Products," 17th Yearbook, National Society for the Study of Education, Part 2, p. 16.

millionths of an inch? In each instance, however, these things existed, even though they had not up to that time been measured with any degree of reliability. In the intervening years very positive strides have been taken in developing instruments that will measure radioactivity, intelligence, millionths of an inch, as well as many other things. Our present instruments for measuring achievement are crude in comparison with the various electronic devices used in physical measurements, but definite improvements are being made continuously. In the years ahead we shall be able to place more and more faith in the results of such tests. And just as the physical scientist is ever striving to improve the measuring instruments of his profession, so must we in teaching endeavor not only to make better measuring instruments but to make better use of those which are already available. This is not a job for the testing expert alone. It is a vital part of effective teaching.

In this discussion we are primarily interested in the measurement of achievement in the classroom, the shop, or the laboratory or on the job. You will be investigating and learning how to construct instruments for best determining relatively what a student has learned, how well he has achieved, how much he has developed, or to what extent he has changed. You will be learning how to use the measuring devices already available.

These goals are easy to understand. They are not so easy to attain. The point to be remembered is that achievement can be measured if adequate instruments are developed to do the job and if you use these instruments properly. Your responsibility as a teacher is to make or use measuring devices that are as accurate as it is now possible to obtain. This is a challenge that requires a thorough understanding of the progress that has been made to date as well as the problems and difficulties that still remain to be solved.

How Accurate Should Measurements Be?

You have just read that the measuring instruments of the teacher should be as accurate as possible. This statement might be tempered somewhat, in the following manner: The measurement should be as accurate as is necessary in terms of the objectives set forth. As an example of this modification, suppose that someone

asked you for a piece of string "about 3 feet long." Your method of measurement would probably be somewhat different from that you would adopt if you were asked for a piece of string "exactly 3 feet long." The point is that in many instances an approximate figure would be sufficient. In other cases it would have to be more accurate. This can be illustrated further by two examples that have their setting in a teaching situation.

If a student were making a freehand pictorial sketch of an end table, you might tell him, in response to a question, that the length should be about 6 inches on the drawing paper. This might be sufficient for his purpose. If, however, he were making the detailed working drawings of the table, he would want the measurements to be more accurate and made with a scale. In this instance the use determines the degree of accuracy that is necessary. A similar situation would exist if a student's father were to meet you on the street and ask how his boy is getting along in your class. You might answer "fine," or "pretty good," or "not very well." At best this would be an approximation, but it might be sufficient for the purpose. On the other hand, if the father were to come to you at school and want to know in more detail about Joe's accomplishments, then conceivably your appraisal should be more detailed and more accurate. You should be able to show him some specific measurements of the boy's relative achievement. This is not always easy to do in an accurate manner.

Sometimes a very rough measure of achievement will suffice. In most cases this can be gained by brief, objective observation of the student at work. At other times you will want the measurement to be as accurate as possible. Then you look for the most precise instrument that can be obtained. A well-prepared test may help to provide such a measurement.

Another aspect of approximate measurement can be illustrated by considering the thickness of the page you are now reading. Suppose you were asked to measure the thickness. How would you proceed? Many readers would think of a micrometer because that measures in terms of thousandths of an inch. A micrometer would probably solve the problem, but suppose you could not read the scale and did not want to get outside assistance. A simple procedure would be to get enough sheets of paper to make a pile 1 inch

high. By counting the number of sheets, you could determine roughly the thickness of a single sheet (in terms of a fraction of an inch). This measurement would be satisfactory for many purposes. It would not be as accurate as that determined with a micrometer, but it would be much better than a pure guess.

This example illustrates the point that some measuring instruments may be more accurate than others but may also be more difficult to use. Since achievement tests are measuring instruments (crude as they sometimes are), this likewise applies to them. Some tests, like the micrometer, cannot be used by the ordinary teacher because of complex procedures and "scales" which he does not understand or has not learned to use. But just as it is possible to use a ruler to obtain a rough measurement of page thickness, it is also possible for a teacher to learn how to devise instruments which will provide a measure of his students' achievement that is much better than a pure guess.

With further study and application he can learn to make or use tests that are more refined (although the measurement will still be crude in comparison with that obtained with a micrometer). The corollary can be carried one step further by stating that certain aspects of measuring achievement are similar to measuring the exact thickness of this page—sometimes it is a job for the testing expert, the person whose efforts are concentrated in this special field.

Steps in Developing a Measuring Instrument

Another fundamental of measurement can be introduced by referring again to the words and paragraphs you are now reading. Suppose a person has just asked you to measure the printing on this page. Suppose further that you have agreed to do this and have just sat down to comply with the request. Without doubt you would soon be wondering just what property the person wanted measured. If you were to jot down the possibilities on a piece of scratch paper, they might look something like the illustration on the next page.

This might be only a partial list of possible things to measure, but it would probably be sufficient for you to go back to the individual and ask him to state exactly what he wanted you to measure. Let's say he asks you to measure the length of the lines and

the printing ink. This time you are not put off easily. We'll assume that you understand about measuring the length of lines, but not the printing ink. So you insist that he be more specific as to what he wants measured with respect to the ink. He decides that he is interested primarily in measuring the thickness of the ink on this page.

At this stage you may have determined to your satisfaction just what it is that you are supposed to measure (length of lines and thickness of ink). The next step is to do the measuring, and an invariable question will be to determine what instrument to use.

Things on this page that might be measured
1 Length of lines
2 Space between lines
3 Height of the type
4 Width of the type
5 The printing ink
6 Page balance
7 Proportions

With respect to the length of lines you would probably start out by using a rule. You would measure the line and find it to be $4\frac{5}{16}$ inches long. Suppose you then take this figure to your friend in evidence of your progress. He might be satisfied, or he might tell you that your answer is not satisfactory—he wanted the measurement in terms of picas (a standard measuring unit in printing). Your first reaction might be to tell your friend that this should have been explained to you at the outset. He did not do a good job of telling you "what" to measure with respect to length (that is, what scale or unit to use).

Perhaps you give your friend a lecture on the importance of describing *exactly what* he wants done. When that is finished, you still have to determine *how* to get the line measured in terms of picas. In this instance it is relatively easy because there is a standard instrument (line gauge) for measuring such units. Your problem then is to obtain a line gauge from a print shop and measure the lines. You would find them to be 26 picas long.

The second part of your assignment is to find out the thickness

of the ink on the page. In this instance it is clear "what" is to be measured. Measuring the length of the lines was relatively easy after you had determined specifically "what" was to be measured. Now the problem is changed somewhat because you understand clearly what is to be measured, but the means for doing it presents a problem. If you are an amateur physicist or an ingenious individual, you *might* devise an instrument or a method for making the measurement. At best, this would take considerable research and experimentation. Your best bet would be to consult a specialist in physical measurement. He might decide to utilize or devise an instrument based on the principle of light reflection (this is one method that has been used in measuring thin films). The point is that a special means would have to be employed, and this would entail experiences that are not familiar to the ordinary individual.

The situation described above illustrates some important points with respect to measurement of any kind. In the first place it should clarify the two major steps to follow in developing an achievement test or any other measuring instrument.

1. *Determine* exactly *what is to be measured.*
2. *Obtain or construct a measuring instrument that will best do the measuring.*

This fundamental can be summarized by the words "what" and "how." It should be easy to understand what is meant by these two steps. Carrying them out is often a complex matter. We can justifiably say that the remainder of this book is devoted largely to helping you determine "what" to measure and "how best" to measure it.

With these thoughts in mind it might be well to reflect once again on the problems that were met in measuring the length of the lines and thickness of the ink on this page. Several difficulties arose in determining "what" to measure. This was narrowed down to length and thickness. However, before the length could be measured, the "what" had to be described more specifically. When that was accomplished, it became relatively easy to determine the number of picas by using the standard line gauge. In this instance the "what" took a little time, but the "how" was relatively easy. In measuring the thickness, it was simple to determine the "what," but "how" to do it presented some problems.

In measuring human achievement the procedure is exactly the same *but much more complex.* It is still a matter of determining "what" and "how," but it is considerably more difficult to state in precise terms the things that are to be measured and after that to construct measuring instruments that will do the job. Nevertheless, the same basic procedure is followed. By putting the "what" and the "how" into question form they become directly applicable to your teaching and testing in the classroom or shop: "What am I trying to measure and how can I best do the measuring?" *If you as an instructor will do nothing more than conscientiously ask and try to answer these two questions, the examinations you make and use will be bound to improve.*

Comparison of Physical Measurement and Achievement Tests

A ruler is a measuring instrument. An achievement test is also a measuring instrument. Just as some rulers are more accurate than others, some achievement tests are more precise than others. Rulers have been in existence for a very long time. They did not just happen. They had to be developed. Over the years they have been refined and improved in a variety of ways. In comparison with achievement tests rulers are used to measure a very simple property—length. On the other hand, human achievement is a much more complex property, and it is only in recent years that efforts have been concentrated on such measurement.

These few statements may appear digressive, but they are intended to serve a specific purpose—to make you think of an achievement test as a measuring instrument, just as a ruler is a measuring instrument, and also to emphasize the necessity of being careful in trying to carry the comparison any further than this. The logic of this suggestion can be justified by pointing out some significant ways in which these types of measurement differ.

Absolute and Relative Measurement

One important difference can be noted by a brief consideration of *absolute* and *relative* measurement. Absolute measurement answers the question "how much." A single expression given in such terms has definite meaning. It can stand by itself. It means the same in one part of the country as in another. We can say that a

person is so many inches tall, or weighs so many pounds, or runs 100 yards in so many seconds. This has definite meaning in terms of certain prescribed standards.

Every reader will understand the statement "He is 6 feet tall." When we say that Junior received a score of 72 in a history test, however, the meaning is not similarly definite and in terms of well-defined and precise standards. The score will mean little until it is related specifically to the test (the number of items, the number of students, the type of students, the difficulty of the items, etc.). The phrase "6 feet tall" means 72 inches all over the country, but what does the score of 72 on the history test really mean? It has little meaning by itself. Before the test score can be interpreted wisely, it must be related to the test and those who were tested. This is an oversimplified way of saying that it is *relative* in nature.

Another differentiating aspect can be explained by considering the use of "zero." One meaning of the cipher, zero, is that it is the point from which measurements are made. Sometimes this point is chosen arbitrarily, as the zero on a thermometer or the point from which the meridians of longitude are computed. In absolute measurement, however, it is possible to establish zero so that it indicates a total lack of the property being measured. In the case of temperature, it is called "absolute zero" (—273.1° C.), a computed point that is believed to register the total absence of heat.

In the measurement of achievement we have not been able to establish a similar point that registers a total lack of the property. We do not have an "absolute" starting point upon which to base measuring scales. This can be illustrated further by comparing the measurement of length with that of achievement.

We can say with assurance that 4 inches is twice as much as 2 inches because zero on a ruler indicates a total lack of length. However, we cannot say that, since John gets a test score of 60 and Bill gets a score of 30, John has achieved twice as much as Bill, for zero on the test would not indicate a total lack of achievement. If a particular student had received a score of zero, it would only mean that he failed to get any of the items right. It would not mean that he had achieved absolutely nothing during the period covered by the test. Bill answered 30 items correctly, and Joe's