Measurement and Evaluation in Human Performance

James R. Morrow, Jr.
Allen W. Jackson
James G. Disch
Dale P. Mood

Measurement and Evaluation in Human Performance

James R. Morrow, Jr., PhD
University of North Texas

Allen W. Jackson, EdD
University of North Texas

James G. Disch, PED
Rice University

Dale P. Mood, PhD
University of Colorado

Human Kinetics

Library of Congress Cataloging-in-Publication Data

Measurement and evaluation in human performance / James R. Morrow, Jr. . . . [et al.].
p. cm.
Includes 1 Macintosh or IBM PC compatible disk.
Includes bibliographical references and index.
ISBN 0-87322-731-X (IBM version). -- ISBN 0-87322-961-4 (Macintosh version)
1. Physical fitness--Measurement. 2. Physical fitness--Testing.
I. Morrow, James R., 1947- .
QP301.M3755 1995
613.7'028'7--dc20 94-44194
CIP

ISBN: 0-87322-731-X (IBM/PC-compatible version); 0-87322-961-4 (Macintosh version)

MYSTAT is a product of SYSTAT, Inc., 1800 Sherman Avenue, Evanston, Illinois.

Permission notices for material printed in this book from other sources can be found on page vi.

Acquisitions Editor: Richard Frey, PhD
Developmental Editor: Christine Drews
Assistant Editors: John Wentworth, Karen Bojda, and Anna Curry
Copyeditor: Jay Thomas
Proofreader: Sue Fetters
Indexer: Jacqueline Brownstein
Typesetter and Text Layout: Kathy Boudreau-Fuoss
Text Designer: Judy Henderson
Photo Editor: Cliff Maier
Cover Designer: Jack Davis
Photographer (cover): John Kelly
Printer: Braun-Brumfield

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Human Kinetics
P.O. Box 5076, Champaign, IL 61825-5076
1-800-747-4457

Canada: Human Kinetics, Box 24040, Windsor, ON N8Y 4Y9
1-800-465-7301 (in Canada only)

Europe: Human Kinetics, P.O. Box IW14, Leeds LS16 6TR, England
(44) 532 781708

Australia: Human Kinetics, 2 Ingrid Street, Clapham 5062, South Australia
(08) 371 3755

New Zealand: Human Kinetics, P.O. Box 105-231, Auckland 1
(09) 309 2259

Preface

Testing and measurement are central to the field of human performance. In teaching a class, conducting a written or performance test, completing a fitness evaluation, or selecting team membership, each of us makes evaluative decisions daily and desires to make good decisions. The basic concepts of testing include measuring and evaluating results and making fundamental decisions. To make accurate decisions, it is imperative that you have a firm foundation in basic measurement concepts. This book provides a wealth of measurement information that applies to all areas of human performance as well as to the health sciences.

To make appropriate decisions a person must be able to assemble information, sort it, evaluate it, and draw conclusions. Performing these processes requires understanding the concepts of reliability, validity, and objectivity. We will provide you with the information and techniques you need to determine whether information is reliable, valid, and objective for decision making in a variety of academic and nonacademic settings. We present this information in three main sections. Part I, "Basic Tools in Measurement and Evaluation," gives essential tools to organize, assimilate, and reduce information for analysis. Part II, "Reliability, Validity, and Grading," presents statistical information in a framework to help judge the quality of data. Essentially, this information deals with two questions: Are the data being obtained accurately? and, Are the data associated truthfully with what you think is being measured? These questions will be answered in a variety of situations. Part III, "Applications of Measurement and Evaluation," provides applications of basic statistical techniques, reliability, and validity to practical problems in the field of human performance and movement.

Often students are apprehensive about the mathematical skills necessary for performing well in measurement and evaluation. Although it is true that a strong mathematical background would provide the student with an advantage, we present our material in such a way that minimal mathematical expertise is expected or required. Moreover, the use of computers makes the mathematics relatively simple. You should bear in mind, however, that you must understand the theory if you are to make effective decisions. We therefore present the theory and the skills that you will need to use when making decisions based on testing, measurement, and evaluation.

In this text we also attempt to provide a unique presentation of material and to include controversial and issue-oriented topics. "Unique presentation" comprises the extensive use of summary tables, descriptive figures and graphs, ***highlighted words and sentences***, a running glossary in the margins of the pages, and Mastery Items (MIs), which both emphasize key points made in the text and require you to apply the principles you have learned to solve problems.

It is important that you complete all of the Mastery Items (MIs) presented to fully understand the material in the chapters. You should take full advantage of your personal copy of MYSTAT and not only conduct the tasks assigned within the text but also create and analyze your own real and hypothetical data sets. Our desire is that you will become a competent MYSTAT user with a good understanding of basic statistical concepts.

Our goal is to present a very visual text that will help you focus on the important concepts and serve as a resource in the learning experience. The measurement and evaluation process in physical education, athletics, and exercise and sport science is a complex one in which conventional knowledge and practice do not always mirror the findings and conclusions of scientific research. Our goal is for you to have an understanding of the measurement and evaluation processes that are scientifically sound and essential for making valid decisions.

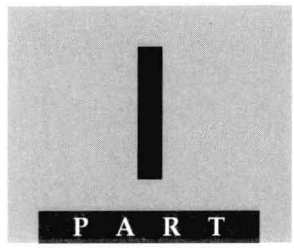
Each of us authors has taught the course for which this book is intended, and we hope our insight and experiences will help you. We've known each other for over 20 years and believe that we share common beliefs and understandings about the importance of the measurement and evaluation process. We know each other as coauthors and as friends. We hope you enjoy and learn from our presentation.

This text could not have been completed without much guidance, many suggestions, and a great deal of encouragement from the professionals at Human Kinetics. Our association with Human Kinetics has been most rewarding. Rainer and Julie Martens can be proud of their organization. We particularly acknowledge Rick Frey and Christine Drews, who spent a great deal of time with us. Judy Henderson and Kathy Boudreau-Fuoss spent much time and effort designing the layout and presentation, and John Wentworth helped us to tie things together at the end. Their efforts are greatly appreciated. We thank the blind reviewers who provided us with many helpful suggestions. We value the measurement and evaluation professionals from whom we have learned much. These (our mentors, friends, and students) include ASJ, ATS, BAM, CHS, DJH, GVG, HHM, JAS, JEF, JLW, KDH, LDH, LSF, MAL, MEC, MJS, MJS, MSB, MTM, RGF, RWS, SSS, TAB, TMW, VWS, and WBE. Lastly, we acknowledge our families, who tolerated our many hours before the computer.

Contents

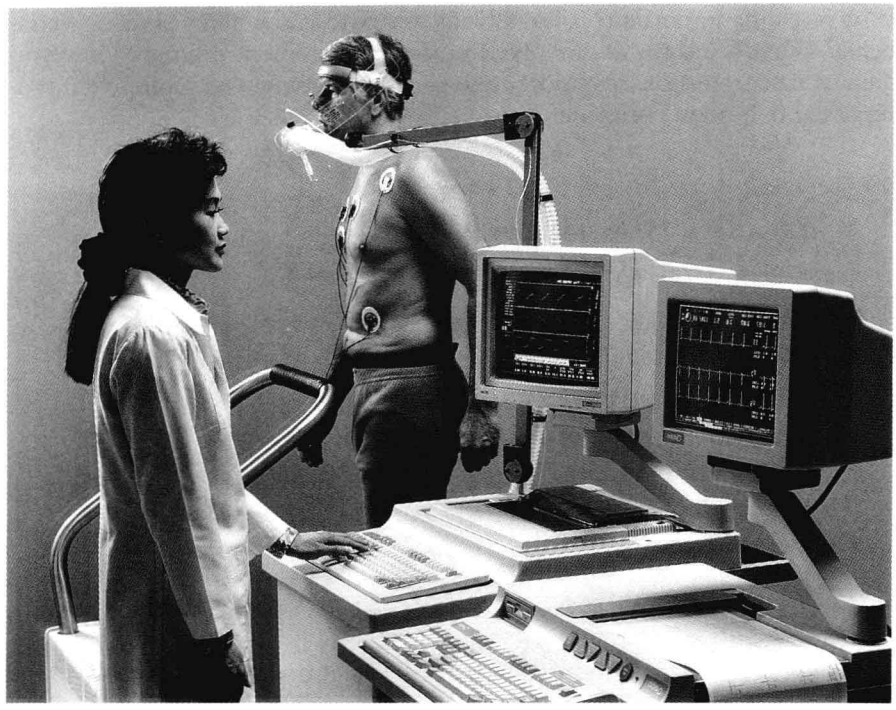
Credits

Page Number	
3	Photo courtesy of Quinton Instrument Company
15, 103, 242, 245, 257, 258	Photos by Karen Maier
17	Photo courtesy of University of Illinois Archives
23	The Prudential FITNESSGRAM used by permission of The Cooper Institute for Aerobics Research
25	Photo by Lorri Bettenga
27	Photo courtesy of Johnson & Johnson Health Management, Inc.
42	Table 3.2 from *A First Course in Statistics*, © 1942 by E.F. Lindquist. Reprinted by permission of Houghton Mifflin Company
56, 332	Photos courtesy of *The Daily Illini*, Champaign, Illinois
72	Photo by James Wentworth
77, 91, 99, 123, 148, 265, 269	Photos from *Physical Education for Children: Concepts Into Practice* (pp. 88, 17, 43, 67, 61, 35, 148) by J.R. Thomas, A.M. Lee, K.T. Thomas, 1988, Champaign, IL: Human Kinetics. Copyright 1988 by Jerry R. Thomas, Katherine T. Thomas, and Amelia M. Lee.
96, 98	Tables 4.11 and 4.14 from A Golf Skills Test Battery for College Males and Females, *Research Quarterly for Exercise and Sport*, Vol. 58, pp. 72-76. Reproduced with permission from the American Alliance for Health, Physical Education, Recreation and Dance, Reston, VA.
95, 208	Photos by David E. Martin
106, 160, 172	Photos by Cliff Maier
110	Photo by Mary E. Messenger
130	Photo courtesy of the Champaign Park District, Champaign, IL
153	Photo by Michael Riley, California State University, Fullerton
218	Photo courtesy of Linda K. Gilkey. From *Advanced Fitness Assessment and Exercise Prescription* (2nd ed., p. 144) by V.H. Heyward, 1991, Champaign, IL: Human Kinetics. Copyright 1991 by Vivian H. Heyward.
220	Photo courtesy of British Indicators, West Sussex
229	Photo courtesy of MedX Corporation
288	Table 10.3 reprinted with permission from the American Alliance for Health, Physical Education, Recreation and Dance, Reston, VA.
303	Photo by Tom King. From *Hit It! Your Complete Guide to Water Skiing* (p. 122) by B. Kistler, 1988, Champaign, IL: Human Kinetics. Copyright 1988 by Bruce Kistler.



Basic Tools in Measurement and Evaluation

We all want to make good decisions. In Part 1 you will be introduced to concepts of measurement and evaluation and their importance in decision making. These concepts are the foundation for your study throughout the remainder of the book. Chapter 1 presents an overview of the scope and use of measurement in human performance. Chapter 2 provides you with a description of computer applications in human performance, with specific consideration given to applications of measurement, testing, and evaluation. Most importantly, this chapter introduces you to MYSTAT, your microcomputer tool for conducting many of the exercises in the remainder of your book and working with the statistics in chapter 3. Chapter 3, on statistics, is extensive, but the mathematics required for successful understanding of the concepts is basic, requiring no more than a background of high school mathematics.



Chapter 1

Introduction to Tests and Measurements

Key Terms

affective domain, p. 10
cognitive domain, p. 10
criterion-referenced standard, p. 5
evaluation, p. 4
formative evaluation, p. 6
measurement, p. 4
norm-referenced standard, p. 5
psychomotor domain, p. 10
summative evaluation, p. 6
taxonomy, p. 10

Why is testing important? Is it really necessary to know many statistical concepts? What decisions are involved in the measurement process? How you answer these questions is important to your development as a competent professional in human performance.

After studying this chapter you will be able to

- define the terms *test*, *measurement*, and *evaluation*,
- differentiate norm- and criterion-referenced standards,
- differentiate formative and summative evaluation,
- discuss the importance of measurement and evaluation processes,
- identify the purposes of measurement and evaluation,
- identify the importance of objectives in the decision-making process, and
- differentiate the cognitive, psychomotor, and affective domains as they relate to human performance.

We all gather data before making decisions. It makes no difference if the decision-making process occurs in education or in other pursuits. For example, we gather information for student grades, research projects, and fitness evaluation. Likewise, we gather data about the weather before venturing out for a morning run. Depending on the data we obtain (e.g., rain, warm, dark, cold), we modify our behavior and act in some fashion. Before purchasing a stock for investment, we gather data on the company's history, leadership, earnings, and goals. All of these are examples of testing and measuring. In each case, making the best possible decision is based on collecting data.

The type of course you are embarking upon has historically been called "tests and measurements." Whereas some students refer to it as "statistics," that does not accurately describe what the course is about. Some basic statistical concepts will be presented in chapter 3; however, the statistical and mathematical knowledge necessary for testing and measurement is not extensive. On the other hand, every chapter in this text focuses in some way on the issues of reliability and validity. It is extremely important to make decisions in a valid (truthful) manner. ***Making effective decisions depends on first obtaining relevant information.*** This is where testing and measurement enter the picture.

THE NATURE OF MEASUREMENT AND EVALUATION

measurement—The act of assessing (e.g., assessing a knowledge or psychomotor test score or one's attitude toward physical activity).

The terms we use in measurement and evaluation have very specific meanings. Measurement, test, and evaluation refer to specific elements of the decision process. Although the three terms are related, each has a distinct meaning and should be used correctly. **Measurement** is the act of assessing. Usually this results in assigning a number to the character of whatever is assessed. A **test** is an instrument or tool used to make the particular measurement. This tool may be written, oral, a mechanical device, or another variation. **Evaluation** is a statement of quality, goodness, merit, value, or worthiness about what has been assessed. Evaluation implies decision making.

You can measure a person's maximal oxygen uptake ($\dot{V}O_2$max, a measure of aerobic capacity) in several ways. You might have someone perform a maximal run on a treadmill and collect and analyze expired gases. You might collect expired gases from a maximal cycle ergometer protocol. A subject might perform either a submaximal treadmill exercise or a cycle exercise and then you might predict $\dot{V}O_2$max from heart rate and/or work load. You might measure the distance a person runs in 12 minutes or the time it takes to complete a 1.5-mile run. Each of these tools results in a number, such as percent O_2 and CO_2, heart rate, minutes, or yards. Having assessed $\dot{V}O_2$max with one of these tools does not mean that you have evaluated it. ***Obtaining and reporting data have little meaning unless you reference the data to something.*** This is where evaluation enters the process.

test—An instrument or tool used to make a particular measurement.

evaluation—A dynamic decision-making process that places a value judgment on the quality of what has been measured (e.g., a test score or physical performance).

Assume that you test someone's $\dot{V}O_2$max. Further, assume that she has no knowledge of what the $\dot{V}O_2$max value means. Certainly, the subject might be aware that the treadmill test is used to measure fitness. However, the first question most people ask after completing some measurement is How did I do? or How does it look? To simply report "Your $\dot{V}O_2$max was 30 ml · kg^{-1} · min^{-1}" says little. You need to provide an evaluation. ***An evaluative statement about how good the performance was introduces the element of merit, or quality.***

Mastery Item 1.1

A physical education teacher records the number of sit-ups that a student completes in 1 minute. Differentiate between the test, measurement, and evaluation characteristics reflected in this activity.

Norm- and Criterion-Referenced Standards

To make an evaluative decision, you must have a reference perspective. ***You can make evaluative decisions from either norm-referenced (normative) or criterion-referenced standards.*** An evaluative decision based on a **norm-referenced standard** means that you report how well a performance compares with that of others (perhaps of people of the same gender, age, or class). Thus, you might report that a $\dot{V}O_2$max of 30 is relatively poor for someone's age and gender. Conversely, you might simply report a person's performance relative to a criterion that you would like them to achieve. Assume that the $\dot{V}O_2$max measured was on someone who had had a heart attack. The physician may be interested in whether the patient can achieve a $\dot{V}O_2$max of 25 ml · kg^{-1} · min^{-1}. This is a case of a **criterion-referenced standard**. You have little or no interest in how someone compares with others; the comparison is with the standard, or criterion. The criterion often is initially set based on norm-referenced data and the best judgment of experts in the content area.

norm-referenced standard—A level of achievement relative to a clearly defined subgroup, such as all women or women your age.

criterion-referenced standard—A specific, predetermined level of achievement.

Youth fitness testing has changed greatly in the last decade. Fitness scores used to be norm-referenced, that is, relative to a child's classmates, by age and gender. Many youth fitness tests now are criterion-referenced. Table 1.1 provides an example of the differential interpretation of norm- and criterion-referenced standards for a 10-year-old boy who performed 30 sit-ups. Note that his score of 30 meets the minimum criterion for the FITNESSGRAM but fails to meet the criterion-referenced standard for the American Alliance for Health, Physical Education, Recreation and Dance's (AAHPERD) Physical Best. Table 1.1 also presents the corresponding norm-referenced standards from the AAHPERD *Health-Related Physical Fitness Manual* (1980). A criterion score

of 30 represents the 35th percentile on the AAHPERD Health-Related Fitness Test. A criterion score of 34 represents the 50th percentile.

Table 1.1 Criterion- and Norm-Referenced Standards for 10-Year-Old Boy Who Performed 30 Sit-Ups

Fitness test battery	Criterion test score	AAHPERD Health-Related Fitness Test percentile
FITNESSGRAM health standard	30 sit-ups	35th
Physical Best health fitness standard	34 sit-ups	50th

FITNESSGRAM data from FITNESSGRAM (1987). Dallas: Institute for Aerobics Research.

Physical Best data from AAHPERD Physical Best Program (1988). Reston, VA: AAHPERD.

AAHPERD health-related data from AAHPERD Health-Related Fitness Test (1980). Reston, VA: AAHPERD.

Mastery Item 1.2

Are the following measures evaluated from a norm-referenced or a criterion-referenced perspective?

a. Blood pressure
b. Fitness level
c. Blood cholesterol
d. A written driver's license examination
e. Performance in a college class

Mastery Item 1.3

Think of other examples of norm- and criterion-referenced evaluative comparisons.

Formative and Summative Evaluation

formative evaluation—A judgment conducted during an instruction or training program.

summative evaluation—A final, comprehensive judgment conducted near the end of an instruction or training program.

Evaluations occur in two perspectives, formative and summative. **Formative evaluations** are initial or intermediate evaluations, such as the administration of a pretest and the subsequent evaluation of its scores. Formative evaluation should occur throughout the instructional, training, or research process. Ongoing measurement, evaluation, and feedback are essential to the achievement of the goals in a program in human performance. These ongoing evaluations need not involve formal testing; simple observation and feedback sequences between the student or participant and the instructor or leader are adequate. **Summative evaluations** are final evaluations that typically come at the end of an instructional or training unit. You, as a student in this course, are interested in the summative evaluation—the grade—you will receive at the end of the semester. The world-class athlete's summative evaluation might occur during the Olympic Games, with the winning or losing of a medal.

The difference between formative and summative evaluations might seem to be merely the difference in timing of their data collection; however, it is the

actual use of the data collected that distinguishes the evaluation as formative or summative. Thus, the same data can be used for formative and summative evaluations.

A weight-loss or weight-control program provides a simple and useful example for applying formative and summative evaluations. Assume that you have measured a participant's body weight and percent fat. Your formative evaluation indicates that he has a percent fat of 30% and needs to lose 10 pounds to achieve a desired percent fat of 25%. You establish a diet and exercise program designed to produce a weight loss of 1 pound per week for 10 weeks. Each week you weigh the participant and measure his percent fat, and give him feedback on the formative evaluations you are conducting. The participant knows the amount of progress or lack of progress that is occurring each week. At the end of the 10-week program, you measure his body weight and percent fat and conduct a simple summative evaluation. Were the weight-loss and percent-fat goals achieved at the end of the program?

Mastery Item 1.4

Develop a scenario similar to the one above that is designed for a participant who wishes to improve abdominal strength and endurance and that involves formative and summative evaluations.

PURPOSES OF MEASUREMENT, TESTING, AND EVALUATION

Prospective teachers and researchers in human performance must understand measurement, testing, and evaluation. We make evaluative decisions on a daily basis in human performance. Literally, our students, clients, and colleagues ask us what tools are best and how to interpret and evaluate performance and measurements. These concepts may be the most important that you will study. Related evaluation concepts are objectivity, reliability, relevance, and validity. These terms will be discussed in greater detail in later chapters.

There are many ways to use the evaluative process in human performance teaching and research. For instance, consider the issue of *accountability*. Your employer might hold you accountable for a project. Tests, measurement, and the evaluation process are used to show whether you are accountable. Obviously, you want the evaluation to accurately reflect the results of your work—assuming that you did a good job! Certainly, if you enter the teaching profession, you will hold your students accountable for the content of the courses you teach. Likewise, your students should hold you accountable for preparing the best possible tests for evaluating their class performance.

Mastery Item 1.5

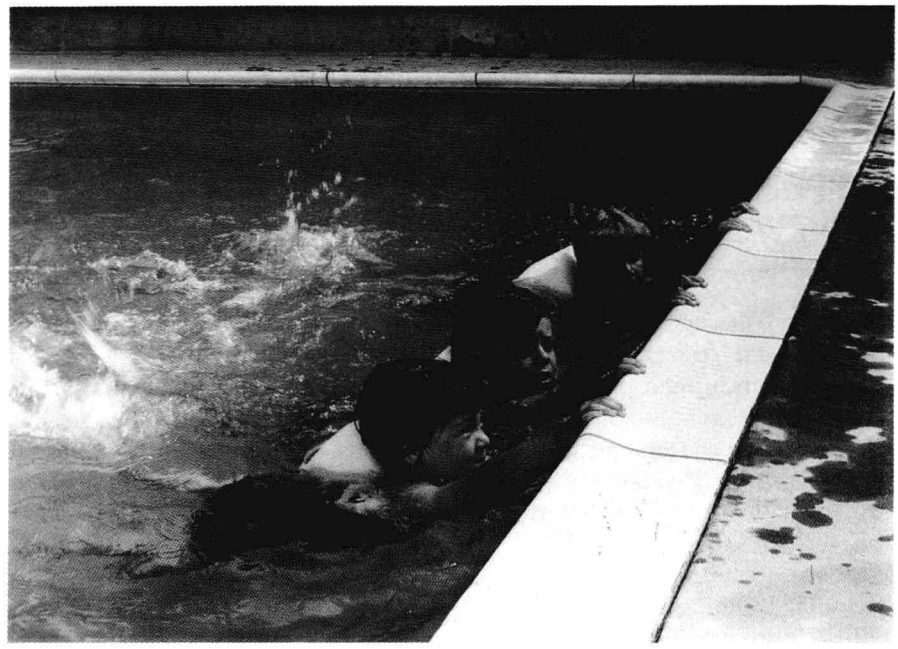
Assume you are a fitness class instructor. How would you determine if your program is effective?

As you will discover during your course of study, you need considerable knowledge and skill to conduct correct and effective measurement and

evaluation. As with any academic or professional effort, ***it is important to have a thorough understanding of the purposes for executing a measurement and evaluation process***. There are six general purposes—*placement*, *diagnosis*, *prediction*, *motivation*, *achievement*, and *program evaluation*.

Placement

An initial test and evaluation allows a professional to group students into instructional or training groups according to their abilities. In some cases, instruction, training, and learning in human performance can be facilitated by grouping together participants in this manner. All participants in a group can then have an appropriate starting point and can all improve at a fairly consistent rate. For example, it is difficult to teach a swimming class if half the students are nonswimmers and the others are members of the swim team.



An initial test and evaluation allows a professional to group students according to their abilities. Reprinted from *Teaching Young Swimmers Manual* with permission from the YMCA of the USA, 101 N. Wacker Drive, Chicago, IL 60606.

Diagnosis

Evaluation of the test results is often used to determine weaknesses or deficiencies in students, medical patients, athletes, and fitness program participants. Cardiologists may administer treadmill stress tests to obtain exercise electrocardiograms for cardiac patients to diagnose the possible presence and magnitude of coronary heart disease. A football coach may have his team complete a series of physical performance tests to determine team and individual conditioning needs.