



# Educational and Psychological Measurement

George K. Cunningham

# EDUCATIONAL and PSYCHOLOGICAL MEASUREMENT

---

**George K. Cunningham**

*University of Louisville*

**MACMILLAN PUBLISHING COMPANY  
NEW YORK**

**Collier Macmillan Publishers  
London**

Copyright © 1986, George K. Cunningham.

Printed in the United States of America

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Macmillan Publishing Company  
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

**Library of Congress Cataloging in Publication Data**

Cunningham, George K.  
Educational and psychological measurement.

Bibliography: p.  
Includes index.

1. Educational tests and measurements. I. Title.  
LB3051.C86 1986 371.2'6 85-7125  
ISBN 0-02-326330-X

Printing: 1 2 3 4 5 6 7 8

Year: 6 7 8 9 0 1 2 3 4 5

**ISBN 0-02-326330-X**

# EDUCATIONAL and PSYCHOLOGICAL MEASUREMENT

***To Nancy and Sarah***

# PREFACE

---

This textbook is intended for use in measurement and evaluation courses that fall between those offered to prospective classroom teachers, for whom test-construction techniques are emphasized, and those offered to graduate students, for whom the focus is psychometric theory. Courses for which this textbook would be appropriate are offered to upper-division undergraduate and graduate students. Enrolled in such courses would be students who are majoring in elementary or secondary education, educational psychology, special education, educational administration, counseling, or any one of several other education majors. In addition, it can be used in measurement courses taught in psychology departments because it includes all of the topics included in such courses, along with content more closely related to education.

The book covers those topics that one would expect to find: derived scores, reliability, validity, test construction, standardized tests, mental ability testing, and personality assessment. Material also is covered that is not usually found in textbooks intended for this audience. For instance, an entire chapter is devoted to computer applications in measurement. Although references to computers and examples of their application to measurement can be found throughout this book, the chapter on computer applications is intended to provide the measurement student with more specific information about the ways computers are used in the field of measurement and evaluation. Information is included on how to get started with a computer, methods of entering data, and data analysis options. Specific information about such topics as item analysis, test banking, tailor-made tests, and computerized standardized test scoring and interpretations can also be found.

Another important advance in the measurement and evaluation field is the application of latent-trait theory to practical measurement problems. Presently the most important application of these techniques is in improving sophisticated standardized tests. In addition to the possibility of increasingly precise and useful methods of interpreting scores, these techniques can be used to develop test banks and create tailor-made tests. One major stumbling block in the application of this theory is its mathematical complexity and consequent inaccessibility to those with a limited background in math. In Chapter 8 there is a readable but detailed explanation of latent-trait theory that even students without a strong mathematical back-

ground can grasp. Although there is not sufficient detail for students to begin applying latent-trait theory, there is an adequate amount of coverage to permit them to understand the increasing number of references and examples of its use that they are likely to encounter.

Another important area of study associated with educational measurement is criterion-referenced testing. It is a very attractive approach to assessment and has seen widespread adoption by school systems across the country. In general it has been accepted in an uncritical fashion and in many cases it has been oversold. Too many of its practitioners have a poor understanding of the correct methods for its implementation. As a result, the types of criterion-referenced tests that are widely accepted in the public school sector are often quite different from those described by experts on criterion-referenced testing. Even though the experts are moving in the direction of more sophisticated descriptions of learning objectives, such as domain referencing, most existing criterion-referenced tests are better described as objective-referenced tests, which are considered to be an inferior approach to evaluation. If criterion-referenced tests are to be used on an increasingly widespread basis, then educators are going to need to become much more knowledgeable about how to use them. Chapter 10, "Criterion-Referenced Testing," provides a description of what is currently known about this approach, the methods of implementation, and the problems associated with its use.

Minimum competency testing is a topic that has received a great deal of attention over the last several years. It is a measurement-oriented response to perceived educational failures. Its historical development, legal implications, strengths, weaknesses, successes, and failures are outlined in Chapter 11. The reader will find the most current information on this controversial topic in this book.

The preparation of material on ability testing presents one of the most difficult tasks for the author of a measurement textbook. Unlike most other aspects of measurement, where the students arrive in class knowing little about the subject, most students already possess a modicum of information on this topic. Although some of it may be accurate, much of it may consist of strongly held prejudices. The textbook writer and the instructor not only are faced with the task of teaching the basic concepts, to permit the student to gain an understanding of ability testing, but they also must disabuse students of their existing misconceptions and deal with their deeply held prejudices. Exacerbating this problem is the emotional aura surrounding the topic. Ability testing is not a subject about which it is easy to remain neutral. Our society harbors radically divergent views on the subject, ranging from those who would put an end to all such assessment to those who view ability tests as a justification for the stratification of our society along racial or ethnic lines. Many of these extreme views are in conflict with the known facts about such testing.

Because most of the disagreement surrounding this topic stems from definitional confusion, the definitions are clarified. The intention is to treat controversial aspects of ability testing in an evenhanded, clear, and easily understood fashion. Each side of the issues is presented, and the reader is placed in a position where he or she can draw conclusions based on facts.

Textbooks for measurement and evaluation courses are filled with methods and techniques for performing statistical operations and obtaining scores and coefficients. The reader is usually not given much information concerning why one method of

technique is selected over another. The tone instead tends either to be nonjudgmental, or one of several methods is recommended without explanation. Often there is no single clearly acceptable technique, and different experts may have their own ideas about which should be used. In this book, rationales are provided to assist the reader in deciding which method or technique should be used.

The field of measurement and evaluation is filled with issues and controversies. These range from minor technical questions about selecting appropriate statistical techniques to some of the most fundamental philosophical questions in Western science, such as the relative importance of the influence of heredity and environment on intelligence. Textbooks sometimes try to avoid these issues by presenting both sides of an argument without taking a position. The reader will find this textbook to be different in that respect. Although both sides of controversies are presented, this book takes a position on the major issues and provides a rationale for the student to help him or her understand what is really going on. Furthermore, the controversies are presented in their historical and social context. This avoidance of equivocation and emphasis on context gives the student a better opportunity to understand why things are done the way they are in measurement to enhance interest and understanding.

I would like to express my appreciation to Norm Gronlund, Daniel Mueller, and Jerome Kapes for their invaluable comments and suggestions on the earlier drafts of the book. I would also like to thank my editors at Macmillan: Lloyd Chilton, for his support and assistance in bringing this project to fruition and Wendy Polhemus, for her help in the production of the book. Finally, I wish to thank Becki McGinty for the outstanding job she did in supervising the typing of the enumerable draft versions of the manuscript.

G. K. C.

**EDUCATIONAL and  
PSYCHOLOGICAL  
MEASUREMENT**

# CONTENTS

---

## 1

### Introduction to Measurement

1

Measurement and Science	2	The Purpose of Measurement	2
Historical Development	3	<i>The Practical Dimension</i>	3
		<i>The Theoretical Dimension</i>	4
Opposition to Testing	8	<i>Theoretical Bases for Defending and Criticizing Testing</i>	9
<i>Causes of the Opposition to Testing</i>	9	<i>The Need for Tests</i>	10
A Theoretical Approach to Measurement	11	<i>Measurement As a Physical Characteristic</i>	12
<i>Measurement Error</i>	12	<i>True Scores and Measurement</i>	12
		<i>Constructs in Measurement</i>	13
<i>Values</i>	15	The Structure of This Textbook	16

## 2

### Statistics

18

Introduction	19	Math Phobia and Statistical Anxiety	19
Attaching Numbers to Traits	20	Assumptions	20
		Scaling	21
<i>Types of Scales</i>	21	<i>Assigning the Appropriate Scale to Data</i>	23
		<i>Controversies in Scaling</i>	24
Types of Statistics	26	<i>Descriptive Statistics</i>	26
		<i>Inferential Statistics</i>	26
Using Statistical Methods in Measurement	27	<i>Frequency Distributions and Graphs</i>	27
<i>Central Tendency</i>	31	<i>Variability</i>	35
		<i>The Normal Curve</i>	41
		<i>Correlation</i>	48

# 3

## Derived Scores

62

Introduction	63	Grade and Age Equivalents	63	<i>Grade Equivalents</i>	63
<i>Age Equivalents</i>	66	Percentile Norms	67	<i>Computation of Percentiles</i>	67
<i>Using Percentiles</i>	67	Standard Scores	70	<i>The Small z-score</i>	70
<i>The Normalized T-Score</i>	71	<i>Other Standard Scores</i>	72	<i>The T-Score</i>	70
				<i>Interpreting Standard Scores</i>	76

# 4

## Computer Applications to Measurement

79

Introduction	80	Using a Computer to Perform Statistical Analyses	80
<i>Attitudes Toward Computers</i>	81	<i>Getting Started</i>	81
<i>Preparing to Do the Data Analysis</i>	84	<i>Types of Computers</i>	81
		<i>Entering Data</i>	82
		<i>Telling the Computer What You Want It to Do</i>	87
Using a Computer to Perform Item Analyses	93		
Test Banking and Tailor-Made Tests	94	<i>Test Banks</i>	94
		<i>Tailor-Made Tests</i>	94
Automated Test Scoring and Interpretation	95		
Future Developments in the Use of Computers in Measurement	97		

# 5

## Reliability

99

How Reliability and Validity Are Related	100	<i>Measuring More Than One Construct</i>	101
Defining Reliability	101	<i>Reliability and Measurement Error</i>	101
<i>Technical Definitions of Reliability</i>	102	Parallel Forms Reliability	102
<i>Factors Suppressing Parallel Forms Reliability</i>	103	<i>Stability</i>	103
<i>The Stability and Internal Consistency of Parallel Tests</i>	105	<i>Internal Consistency</i>	104
Reliability as Internal Consistency	106	<i>Split-Halves Reliability</i>	106
<i>Intercorrelation Among Items</i>	107	<i>Kuder-Richardson 20 (KR-20)</i>	107
<i>Kuder-Richardson 21 (KR-21)</i>	108		
<i>Determining the Internal Consistency of Tests That Are Not Scored Dichotomously</i>	110		
Test-Retest Reliability	111	Technical Factors Affecting Reliability	112
<i>Quality of Items</i>	112	<i>Test Length</i>	112
<i>Sample Size</i>	117	<i>Variability</i>	114
<i>Physical Conditions</i>	117	<i>Guessing</i>	114
		<i>Scorer Reliability</i>	117
		Standard Error of Measurement (SEM)	118

## 6 Validity 120

Introduction	121	Methods of Assessing Validity	121
Criterion-Related Validity	112	<i>Types of Criterion-Related Validity</i>	112
<i>The Difficulty in Obtaining Criterion Measures</i>	123	<i>Restricted Range</i>	123
<i>Which Is the Predictor and Which Is the Criterion?</i>	124	<i>How Large Should the Validity Coefficient Be?</i>	125
Construct-Related Validity	125	<i>Establishing Construct-Related Validity</i>	126
Content-Related Validity	127	<i>The Legal Implications of Content-Related Validity</i>	128
<i>The Limitations of Content-Related Validity</i>	129	<i>Improving Content-Related Validity</i>	130
Limitations of the Categorical View of Validity	130		
<i>Combining Different Approaches to Validity</i>	131	The Special Case of Face Validity	132

## 7 Classroom Testing 134

Introduction	135	Time Commitment	135	Planning the Test	136
Item Writing	136	Essay Tests	137	<i>The Disadvantages of Essay Tests</i>	137
<i>The Advantages of Essay Tests</i>	138	<i>Improving Essay Tests</i>	138		
<i>Improving the Scoring of Essay Tests</i>	140				
Short-Answer Tests	141	<i>Improving the Short-Answer Test</i>	141		
Objective Tests	142	<i>The Advantages of Objective Tests</i>	142		
<i>The Disadvantages of Objective Tests</i>	143	<i>Matching Tests</i>	144	<i>True-False Tests</i>	145
<i>Multiple-Choice Tests</i>	146	Conclusions	149		

## 8 Improving Achievement Tests 152

Steps in Improving Achievement Tests	153	<i>Content-Related Validity</i>		153
<i>Following the Rules for Good Item Construction</i>	154	<i>Item Analysis</i>	154	<i>Conclusions</i> 158
Latent-Trait Theory	158	<i>The Limitations of Standard Scores</i> 159		
<i>The Advantages of the Latent-Trait Approach</i>	159	<i>The Assumptions Necessary for the Model</i> 160		
<i>The Basic Model</i>	160	<i>Other Models</i>	161	<i>Applications</i> 165 <i>Conclusions</i> 167
Bias in Achievement Tests	169			

# 9

## Assigning Grades

171

Introduction 172      The Need for Grades 172  
Standards Used to Compare Students 173      *The Intuitive Approach* 173  
*Ipsative Methods* 174      *Reference to Perfection* 174      *Criterion-Referenced Grading* 177  
*Norm-Referenced Grading* 178      A Basic Model for Grading 179  
*Deciding How Many of Each Grade to Assign* 179      *Who Should Get Which Grade* 180  
Combining Scores 180      *Why Adding Raw Scores Does Not Work* 181  
*Making Sure That Input for Grades Is Properly Weighted* 181      *Should You Go to This Much Trouble?* 183  
*Factors Affecting Weighting* 184      Pass-Fail Grading 184      Grade Inflation 185

# 10

## Criterion-Referenced Testing

187

Introduction 188      Educational Technology 188      *Historical Antecedents* 188  
*Programmed Instruction* 189      *Behavioral Objectives* 189      *Mastery Learning* 190  
*The Use of Educational Technology* 191      *The Limitations of Educational Psychology* 191  
Criterion-Referenced Testing 192      *The Limitations of Norm-Referenced Tests* 192  
*Implementing Criterion-Referenced Tests* 193      *The Response to Criterion-Referenced Testing* 193  
*The Limitations of Criterion-Referenced Tests* 193      *Unresolved Issues in Criterion-Referenced Testing* 194  
Cut-Off Scores 195      *State Model Methods* 195      *Continuum Model Methods* 196  
Reliability 197      *Decision Consistency Reliability* 197      Validity 199  
Amplified Objectives 200      *Domain-Referenced Testing* 200  
*Implementing Domain-Referenced Testing* 201      *Domain-Referenced Reliability* 202  
*Conclusions About Domain-Referenced Testing* 203

# 11

## Minimum Competency Testing

205

Introduction 206      The Reasons for the Growth of the MCT Movement 206  
*Decline in the Quality of Education* 206  
*Unwillingness to Solve Educational Problems Through a Greater Allocation of Funds* 208  
The Ascendancy of MCT 209      *The Application of Criterion-Referenced Testing Methods* 209  
*The Proliferation of MCT* 210      Legal Issues Surrounding MCT 210

<i>Special Education and MCT</i>	211	The Problems Associated with MCT	212
<i>The Advantages of the Criterion-Referenced Approach</i>	213		
<i>The Disadvantages of Criterion-Referenced Testing</i>	213		
<i>Improving MCT with Norm-Referenced Methods</i>	215		

## 12

### Achievement Testing

218

Introduction	219	Achievement and Aptitude	219	Traditional Views	219
<i>Modern Interpretations</i>	219	<i>Increased Blurring of Distinctions</i>	220		
Standardized Achievement and Classroom Tests	220				
Developments in Achievement Testing	222	<i>Test Batteries</i>	222	<i>Diagnostic Tests</i>	222
<i>Criterion-Referenced Tests</i>	223	Available Standardized Achievement Tests	224		
The Comprehensive Test of Basic Skills (CTBS)	225				
Evaluating the Usefulness of Standardized Achievement Tests	230	<i>Reliability</i>	231		
<i>Validity</i>	231	Using Standardized Achievement Tests to Evaluate School Systems	232		
<i>Excluding Students from Taking the Tests</i>	233	<i>The Diagnostic Use of Standardized Achievement Tests</i>	233		
<i>The Time of Test Administration</i>	233				
Achievement Tests Commonly Used for Individual Educational Assessment	234				
<i>The Wide Range Achievement Test</i>	234	<i>The Peabody Individual Achievement Test</i>	236		
<i>The Woodcock Reading Mastery Test</i>	237	<i>The KeyMath Diagnostic Arithmetic Test</i>	239		

## 13

### Mental Ability Testing

242

Introduction	243	Intelligence	243	<i>Definitions of Intelligence</i>	243
Heredity Versus Environment	249	<i>The Social Scientists' View</i>	250		
<i>The Geneticists' View</i>	251	<i>Should We Study the Effects of Heredity on Behavior?</i>	251		
<i>Heritability Studies</i>	252	<i>Criticisms of Heritability Studies</i>	253		
<i>How Important Are Heritability Indexes</i>	254	<i>Conclusions About Heritability Studies</i>	255		
Differences in Cultural and Ethnic Performance on Mental Ability Tests	256				
<i>Heritability Indexes Among Non-Caucasian Populations</i>	256	<i>Studies of Blood Grouping</i>	256		
<i>Studies of Separate Abilities</i>	256	<i>Conclusions</i>	257		
Sex Differences in Intelligence Test Performance	257	<i>Studies of Sex Differences</i>	257		
Bias in Intelligence Testing	259	<i>Factor Analysis</i>	260	<i>Item Analysis</i>	260
<i>Aptitude and Achievement</i>	260	Culture-Fair Tests	261		

# 14

## Individual Intelligence Tests

263

Introduction 264      The Stanford-Binet Tests 264      *Interpreting the Stanford-Binet* 268  
The Wechsler Tests 269      *Global Intelligence* 269      *Deviation IQs* 270  
*Tests Developed by Wechsler* 270      *Subscales on the WISC-R* 270      *Substituting Mazes for Coding* 272  
*Comparing the WISC-R and WAIS-R* 272      *Comparing the WISC-R and WPPSI* 272  
*Interpreting the Wechsler Tests* 272      *Comparisons Between the Stanford-Binet and Wechsler Tests* 274  
The Kaufman Assessment Battery for Children (K-ABC) 275  
*The Sequential Processing Scale* 275      *The Simultaneous Processing Scale* 275  
*The Mental Processing Composite Scale* 275      *The Achievement Scale* 275      *The Nonverbal Scale* 276  
*The Theoretical Basis for the K-ABC* 276      *Nondiscriminatory Testing* 277  
*Conclusions About the K-ABC* 278  
The System of Multicultural Pluralistic Assessment (SOMPA) 279      *The Medical Model* 279  
*The Social System Model* 279      *The Pluralistic Model* 280      *Conclusions About the SOMPA* 281  
The Use of Individual Intelligence Tests to Determine Special Education Placement 282  
*Disproportionality* 282      *Conclusions About Special Education Placement* 283

# 15

## Group Mental Ability Tests

285

Introduction 286      Group Intelligence Tests 286      *The Army Alpha and Beta* 286  
*The Cognitive Abilities Test (CAT)* 288      *The Test of Cognitive Skills (TCS)* 289  
College Aptitude Tests 290      *The Scholastic Aptitude Test (SAT)* 291  
*The American College Testing Program (ACT)* 291      *The Graduate Record Exam (GRE)* 292  
*The Millers Analogies Test (MAT)* 292  
Controversies Surrounding Tests of College Aptitude 293      *The Decline in SAT Scores* 293  
*The Effect of Coaching on SAT Performance* 294      *Validity* 295  
*Cost/Benefit of College Aptitude Tests* 297      *Are the Tests Really Used?* 297  
Screening Job Applicants 298      Multiple-Aptitude Tests 299  
*The Differential Aptitude Test (DAT)* 300      *The General Aptitude Test Battery (GATB)* 301  
*The Armed Services Vocational Aptitude Battery (ASVAB)* 303  
*Conclusions About the Use of Multiple-Aptitude Tests* 306      *Conclusions* 307

# 16

## Personality Assessment

310

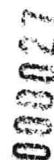
Introduction 311      Learning About Personality from An Individual's Self-Report 312  
*Interviews 312      Questionnaires 313      Objectively Scored Self-Report Instruments 314*  
*Adjustment Scales 315      Interest Inventories 324      Measuring Attitudes 332*  
Learning About Personality Through the Perceptions of Others 337  
*Letters of Recommendation 337      Rating Scales 338      Direct Observations 339*  
*Systematic Observations 339*  
Understanding Personality Through An Examinee's Interpretation of Ambiguous Stimuli 340  
The Rorschach Test 342      The Thematic Apperception Test (TAT) 343  
The Assessment of Personality Through the Interpretation of Drawings 344

# 17

## Selecting Tests, Reporting Results, and the Future of Measurement

346

Introduction 347      Selecting an Assessment Instrument 347      *The Level of Selection 347*  
*The Purposes of Testing 349      Finding Out About Tests and Assessment Instruments 352*  
Finding Standardized Tests 353      Finding Out About the Characteristics of Tests 354  
*Finding Critical Reviews of Tests 354      Evaluating Tests and Assessment Instruments 355*  
*Test Objectives 355      Norm Sampling 356      Types of Derived Scores Provided 357*  
*Reliability and Standard Error of Measurement 357      Validity 358*  
*The Practical Aspects of Instrument Evaluation 359      Reporting Results 361*  
*Conveying Test Results to Colleagues 361      Communicating with Parents 362*  
*Communicating Test Results to the Community 362*  
The Future of Measurement and Evaluation in Education 364  
*Criterion-Referenced Testing 364      MCT 364      Computers in Measurement 365*  
*Latent-Trait Theory 366      Conclusions About the Future of Testing 366*



**APPENDIX: 369**

**BIBLIOGRAPHY 383**

**NAME INDEX 395**

**SUBJECT INDEX 399**

# Introduction to Measurement

---

---

## OVERVIEW

The purpose of this first chapter is to introduce the reader to the basic assumptions upon which this textbook is predicated. A brief history of the development of the measurement field is provided; it is divided into those developments that occurred as a result of the need to solve practical problems and those that have a more theoretical basis. Opposition to testing and its causes, along with the reasons why testing is needed, are included. Finally, a rational basis for measurement, which depends on the importance of constructs, is presented.

## OBJECTIVES

From this chapter you will learn about the history of testing and some of its theoretical underpinnings. Specifically, you will learn the following:

- The relationship between measurement and science.
  - The purpose of measurement.
  - The historical developments that led to the establishment of measurement as a field of study.
  - The reason that there is so much opposition to testing.
  - Why tests are needed.
  - Some of the theoretical bases for testing.
-