

Second Edition

# Educational Measurement and Testing

William Wiersma  
Stephen G. Jurs



---

# **Educational Measurement and Testing**

\_\_\_\_ **Second Edition** \_\_\_\_\_

**William Wiersma**

**Stephen G. Jurs**

*The University of Toledo*

*Allyn and Bacon*

*Boston London Sydney Toronto*

*To Dean George E. Dickson—  
administrator, researcher, scholar, teacher, and friend*

---

*Series Editorial Assistant:* Carol Craig  
*Production Administrator:* Annette Joseph  
*Production Coordinator:* Susan Freese  
*Editorial-Production Service:* TKM Productions  
*Cover Administrator:* Linda K. Dickinson  
*Manufacturing Buyer:* Tamara Johnson



Copyright © 1990, 1985 by Allyn and Bacon  
A Division of Simon & Schuster, Inc.  
160 Gould Street  
Needham Heights, Massachusetts 02194

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without the written permission of the copyright owner.

**Library of Congress Cataloging-in-Publication Data**

Wiersma, William.

Educational measurement and testing / William Wiersma, Stephen G.

Jurs. — 2nd ed.

p. cm.

Includes bibliographical references.

ISBN 0-205-12385-6

1. Educational tests and measurements—United States.

2. Examinations—United States—Study guides. I. Jurs, Stephen G.

II. Title.

LB3051.W493 1990

371.2'6—dc20

89-27090  
CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1    94 93 92 91 90

# Preface

---

As with the first edition, this second edition of *Educational Measurement and Testing* has been designed as a textbook for an introductory course, whether the course is offered on the undergraduate or graduate level. The book may be used for the measurement component in an undergraduate course or instructional block that covers a variety of topics, including measurement and/or testing. The emphasis of the book is on the use of tests in everyday school operation. Teachers or others concerned about constructing tests can use the book as a reference.

The organization of the text is somewhat different than that of the earlier edition, and the content has been changed and updated. There are five parts to the text.

Part One is an introduction to measurement and testing and, although it consists of only one chapter, it serves as an introduction to the entire text. Part Two deals with constructing test items using a variety of formats.

Part Three deals with norm-referenced measurement, and Part Four covers criterion-referenced measurement. Traditional educational measurement is, for the most part, based on norm-referenced concepts. Practically every school requires some norm-referenced measurement, at least through the standardized testing program. Teachers tend to use criterion-referenced measurement in their daily classroom instruction. Therefore, we decided to devote a part to each type of measurement. If instructors desire to cover Part Four before Part Three, this should cause no difficulties. There is adequate repetition of topics, such as necessary statistics.

Part Five contains four chapters about important contemporary testing concerns. These concerns are not unique to either norm-referenced or criterion-referenced measurement. This is not an exhaustive treatment of educational testing concerns and issues, but they are those that we feel are most relevant to teachers and other educators in daily school operation.

The emphasis of the text is on the practical application of measurement and testing procedures. Adequate theory is presented for understanding these procedures, and statistical concepts are introduced as they are required. There is no statistical prerequisite for use of this book. Reliability and validity are covered later in this edition than in the previous one, and these topics are covered in separate chapters for both norm-referenced and criterion-referenced measurement.

Several features of the book are designed to enhance its use as a learning device and as a teaching tool. Important points are summarized and displayed prominently throughout. Key terms and concepts introduced in each chapter are highlighted in the text and listed at the end of that chapter, pinpointing major ideas and facilitating review; these terms are then defined in the Glossary of Key Terms and Concepts.

At the end of each chapter is also a set of review items that the reader may use to check understanding of chapter content; answers to the review items are provided in Appendix 2. In addition, exercises are included for all chapters; solutions to selected exercises are given in Appendix 3.

Two glossaries are provided for the reader's convenience. One contains commonly used measurement terms, as mentioned above; the other contains statistical symbols and formulas.

A mastery of the content of the book will provide the reader with a basic understanding of educational measurement and testing. Since the content emphasis is on concepts and procedures used by classroom teachers, prospective teachers or practicing teachers, along with other educators who work in the school setting, will find this book helpful.

### Acknowledgments

We would like to thank Allyn and Bacon's reviewers, who offered helpful comments in the preparation of this book: Anthony Allen, University of Rhode Island; Arlene Brett, University of Miami; William E. Brown, Wright State University; James R. Collins, University of Wyoming; John J. Graening, University of Arkansas; Donald L. Grigsby, University of Alabama at Birmingham; and Ann Snyder, Oklahoma City University.

W. W.  
S. G. J.

# Brief Contents

---

<b>PART ONE</b>	
<b>INTRODUCTION TO MEASUREMENT</b>	<b>1</b>
1    INTRODUCTION	3
 <b>PART TWO</b>	
<b>TEST CONSTRUCTION</b>	<b>25</b>
2    PLANNING THE TEST	27
3    SELECTED-RESPONSE ITEMS	41
4    CONSTRUCTED-RESPONSE ITEMS	65
 <b>PART THREE</b>	
<b>NORM-REFERENCED MEASUREMENT</b>	<b>91</b>
5    NORM GROUPS	93
6    COMPARING SCORES TO NORM GROUPS	117
7    ITEM STATISTICS FOR NORM-REFERENCED TESTS	137
8    RELIABILITY OF NORM-REFERENCED TESTS	155

9	VALIDITY OF NORM-REFERENCED TESTS	183
 <b>PART FOUR</b>		
	<b>CRITERION-REFERENCED MEASUREMENT</b>	<b>209</b>
10	CRITERION-REFERENCED TESTS	211
11	ITEM STATISTICS FOR CRITERION-REFERENCED TESTS	239
12	RELIABILITY OF CRITERION-REFERENCED TESTS	255
13	VALIDITY OF CRITERION-REFERENCED TESTS	271
 <b>PART FIVE</b>		
	<b>ADDITIONAL TESTING CONCERNS</b>	<b>287</b>
14	FACTORS THAT AFFECT TEST SCORES	289
15	THE USE OF STANDARDIZED ACHIEVEMENT TESTS	309
16	TEST BIAS	333
17	HIGH-STAKES TESTS	353
 <b>APPENDIXES</b>		<b>375</b>
 <b>GLOSSARIES</b>		<b>399</b>
 <b>INDEXES</b>		<b>409</b>

# Contents

---

## PREFACE

xv

## PART ONE

### INTRODUCTION TO MEASUREMENT

1

#### 1 INTRODUCTION

3

Educational Measurement in the 1990s 3

*National Trends* 4

*State Trends* 5

*Local School Trends* 5

*Classroom Trends* 6

Definitions of Measurement Terms 7

*Measurement as Quantification* 7

*Assessment and Testing* 8

Levels of Measurement 9

*Nominal Measurement* 11

*Ordinal Measurement* 11

*Interval Measurement* 12

*Ratio Measurement* 12

Norm-Referenced and Criterion-Referenced Measurement 13

*Norm-Referenced Interpretation* 14

*Criterion-Referenced Interpretation* 15

*Some Distinctions between Norm-Referenced and*

*Criterion-Referenced Tests* 16

Comments about Test Terminology 17

Summary 19

Key Terms and Concepts 19

v



Review Items	20
Exercises	22
References	23

## PART TWO TEST CONSTRUCTION

25

### 2 PLANNING THE TEST

27

Purposes of Tests	27
Knowledge and Educational Objectives	29
<i>Objectives and Educational Outcomes</i>	30
<i>Implications for Test Planning</i>	33
Table of Specifications	34
Practical Considerations in Planning Tests	35
<i>Type of Test</i>	36
<i>Test Length</i>	36
<i>Item Formats</i>	37
Summary	37
Key Terms and Concepts	38
Review Items	38
Exercises	40
References	40

### 3 SELECTED-RESPONSE ITEMS

41

Selected-Response Item Formats	42
General Characteristics of Selected-Response Items	42
<i>Reliable Scoring</i>	43
<i>Content Sampling</i>	43
<i>Tendency to Measure Trivia</i>	44
True-False Items	44
Multiple-Choice Items	48
<i>Advantages of Multiple-Choice Testing</i>	53
Matching Items	54
<i>Constructing the Matching Items</i>	55
General Suggestions for Item Writing	59
Summary	60
Key Terms and Concepts	61
Review Items	61
Exercises	63
References	64

<b>4</b>	<b>CONSTRUCTED-RESPONSE ITEMS</b>	<b>65</b>
	Short-Answer Items	65
	<i>Constructing Short-Answer Items</i>	66
	<i>Question Form</i>	66
	<i>Association Form</i>	67
	<i>Completion Form</i>	67
	Essay Items	69
	Essay Items and Learning Outcomes	71
	Characteristics of Essay Items	72
	<i>Distinctions between Objective and Essay Items</i>	72
	<i>Advantages of Essay Items</i>	73
	<i>Disadvantages of Essay Items</i>	75
	Constructing Essay Items	76
	<i>Restricting the Item and the Response</i>	77
	<i>Directing the Student to the Desired Response</i>	79
	<i>Other Considerations</i>	80
	Scoring Essay Items	82
	<i>Approaches to Scoring</i>	82
	<i>Sources of Scoring Inconsistencies</i>	83
	<i>Procedures for Scoring</i>	84
	Summary	85
	Key Terms and Concepts	86
	Review Items	86
	Exercises	88
	References	89
	<b>PART THREE</b>	
	<b>NORM-REFERENCED MEASUREMENT</b>	<b>91</b>
<b>5</b>	<b>NORM GROUPS</b>	<b>93</b>
	Criteria for Norms	94
	<i>Representativeness</i>	94
	<i>Relevance</i>	95
	<i>Recency</i>	96
	Types of Norms	97
	<i>National Norms</i>	97
	<i>Local Norms</i>	98
	<i>Norms for Subgroups</i>	100
	Norms versus Standards	100
	Descriptive Statistics for Norm Groups	101

	<i>Frequency Distributions</i>	102
	<i>Measures of Central Tendency</i>	104
	<i>Measures of Dispersion</i>	107
	<i>Graphing Distributions</i>	110
	Summary	111
	Key Terms and Concepts	112
	Review Items	112
	Exercises	114
	References	115
<b>6</b>	<b>COMPARING SCORES TO NORM GROUPS</b>	<b>117</b>
	Percentiles	117
	Standard Scores	118
	<i>The Standard Normal Distribution</i>	119
	<i>Transformed Standard Scores</i>	122
	Stanines	122
	Normal Curve Equivalents Scores	124
	Comparison of Various Forms of Standard Scores	125
	Grade Equivalent Scores	125
	An Example from a Standardized Test	127
	Summary	132
	Key Terms and Concepts	132
	Review Items	133
	Exercises	135
	References	136
<b>7</b>	<b>ITEM STATISTICS FOR NORM-REFERENCED TESTS</b>	<b>137</b>
	Correlation Coefficient as a Descriptive Statistic	138
	Difficulty Index	143
	Discrimination Index	145
	Item Selection and Revision	146
	Summary	148
	Key Terms and Concepts	149
	Review Items	149
	Exercises	151
	References	153
<b>8</b>	<b>RELIABILITY OF NORM-REFERENCED TESTS</b>	<b>155</b>
	Stability Reliability	156
	<i>Procedure for Estimating Stability Reliability—Test-Retest</i>	157

Equivalence Reliability	158
<i>Procedure for Estimating Equivalence Reliability—Parallel Forms</i>	158
Internal Consistency Reliability	159
<i>Procedures for Estimating Internal Consistency Reliability</i>	159
Effect of Test Length on Reliability	163
<i>Spearman-Brown Procedure for Estimating Total Test Reliability</i>	163
The Size of the Reliability Coefficient	165
Reliability of Difference Scores	166
Quantitative Concepts of Test Scores	168
<i>Reliability and Variance</i>	170
<i>The Standard Error of Measurement</i>	171
Reliability and an Individual Score	171
Ways to Enhance Test Reliability	173
Summary	175
Key Terms and Concepts	177
Review Items	177
Exercises	180
References	182
 9 VALIDITY OF NORM-REFERENCED TESTS	 183
Content Validity	183
<i>Content Validity of Teacher-Constructed Tests</i>	185
<i>Content Validity of Published Tests</i>	186
Criterion Validity	189
<i>Concurrent Validity</i>	190
<i>Predictive Validity</i>	191
Construct Validity	193
<i>Correlational Analysis and Construct Validity</i>	194
The Relationship between Reliability and Validity	197
<i>The Validity Coefficient</i>	199
Ways to Enhance Validity	200
<i>Factors Internal to the Test</i>	200
<i>Factors External to the Test</i>	201
Summary	202
Key Terms and Concepts	203
Review Items	204
Exercises	207
References	208

<b>PART FOUR</b>	
<b>CRITERION-REFERENCED MEASUREMENT</b>	<b>209</b>
<b>10 CRITERION-REFERENCED TESTS</b>	<b>211</b>
Domain Specification	212
<i>Item Forms</i>	213
Objectives	215
<i>Example of an Early Objectives-Based Test</i>	216
<i>Examples of Recent Objectives-Based Tests</i>	217
<i>Test Specifications</i>	221
Teacher-Constructed Criterion-Referenced Tests	225
Mastery Decisions	228
Methods of Setting Standards	229
<i>Professional Judgment</i>	229
<i>Nedelsky Method</i>	230
<i>Angoff Method</i>	231
<i>Contrasting Groups</i>	231
<i>Standards from Norms</i>	233
Summary	234
Key Terms and Concepts	235
Review Items	235
Exercises	237
References	238
<b>11 ITEM STATISTICS FOR CRITERION-REFERENCED TESTS</b>	<b>239</b>
Item-Level Analysis	239
Uses of Item Information	240
<i>Decisions about Students</i>	240
<i>Decisions about Instruction</i>	240
<i>Decisions about Test Items</i>	241
Item Analysis	241
<i>Content Review before Testing</i>	241
<i>Pilot Testing</i>	242
<i>Analysis of Error Patterns</i>	242
Statistical Item Analysis	243
<i>Item Difficulty</i>	243
<i>Item Discrimination</i>	244
Example of an Item Analysis	246
Summary	248
Key Terms and Concepts	249
Review Items	250

Exercises	252	
References	253	
<b>12</b>	<b>RELIABILITY OF CRITERION-REFERENCED TESTS</b>	<b>255</b>
The Term <i>Reliability</i>	256	
The Reliability (Consistency) of Mastery Decisions	257	
<i>Calculating the Reliability Estimate</i>	258	
The Reliability of the Estimate of a Domain Score	259	
<i>Comments about Se</i>	261	
Use of Norm-Referenced Reliability Coefficients	262	
General Comments about Reliability	263	
<i>Increasing Test Reliability</i>	264	
Summary	265	
Key Terms and Concepts	266	
Review Items	266	
Exercises	268	
References	269	
<b>13</b>	<b>VALIDITY OF CRITERION-REFERENCED TESTS</b>	<b>271</b>
Content Validity	272	
Criterion Validity	274	
<i>Correlation of Domain Scores</i>	275	
<i>Prediction of Mastery</i>	276	
<i>Comments on Criterion Validity</i>	278	
Construct Validity	279	
<i>An Example of Construct Validity</i>	279	
<i>Comments on Construct Validity</i>	281	
Summary	282	
Key Terms and Concepts	283	
Review Items	283	
Exercises	285	
References	286	
<b>PART FIVE</b>		
<b>ADDITIONAL TESTING CONCERNS</b>		<b>287</b>
<b>14</b>	<b>FACTORS THAT AFFECT TEST SCORES</b>	<b>289</b>
Personal Factors	289	
<i>Training in Test-Taking Skills</i>	290	

<i>Guessing</i>	290
<i>Changing Answers</i>	294
<i>Positional Preference</i>	294
<i>Penmanship and Spelling</i>	295
<i>Bluffing</i>	296
<i>Reading Difficulty of Responses</i>	296
<i>Anxiety</i>	297
Test Administration Factors	297
<i>The Use of Separate Answer Sheets</i>	298
<i>The Testing Arrangement</i>	298
<i>Group versus Individual Testing</i>	299
<i>Take-Home and Oral Exams</i>	300
Preparing Students to Take Tests	301
<i>Understanding Instructions</i>	301
<i>Answer Sheets</i>	301
<i>Speed</i>	302
<i>Checking</i>	302
<i>Sleep and Food</i>	303
Summary	303
Key Terms and Concepts	303
Review Items	304
Exercises	305
References	306
<b>15</b>	<b>THE USE OF STANDARDIZED ACHIEVEMENT TESTS    309</b>
Standards for Standardized Tests	310
Selecting Standardized Achievement Tests	312
<i>Relevance</i>	312
<i>Technical Adequacy</i>	312
<i>Usability</i>	314
Standardized Achievement Tests	314
The Teacher's Role and Responsibility	315
<i>Prepare the Physical Setting</i>	320
<i>Prepare the Psychological Setting</i>	321
<i>Follow Directions</i>	321
<i>Score Reporting</i>	322
Limitations of Standardized Achievement Tests	322
<i>Testing Time and Disruptions</i>	323
<i>Timing of Test Administration</i>	323
<i>Test Scores—Specific or Global?</i>	324
<i>Accuracy of Test Scores</i>	324
<i>Test Scores in Isolation</i>	324

	<i>Item Content</i>	324	
	<i>Norms versus Standards</i>	325	
	Computerized Testing in the Future	325	
	Summary	326	
	Key Terms and Concepts	327	
	Review Items	328	
	Exercises	330	
	References	331	
<b>16</b>	<b>TEST BIAS</b>		<b>333</b>
	What Is a Biased Test?	334	
	Performance Bias	334	
	<i>Item Bias</i>	336	
	Content Bias	337	
	What Is Being Done about Bias?	338	
	Legal Aspects of Testing	340	
	<i>Tests Used for Screening and Placement</i>	340	
	<i>Truth-in-Testing Laws</i>	342	
	<i>Student Rights</i>	342	
	Ethical Considerations and Acceptable Procedures for Test Use	343	
	<i>Test Use</i>	344	
	<i>Use of Test Results</i>	346	
	<i>Other Considerations</i>	347	
	Summary	348	
	Key Terms and Concepts	349	
	Review Items	349	
	Exercises	351	
	References	352	
<b>17</b>	<b>HIGH-STAKES TESTS</b>		<b>353</b>
	Minimum Competency Testing	353	
	<i>The Meaning of Minimum Competency Testing</i>	354	
	<i>Types of Minimum Competency Tests</i>	355	
	<i>Rationale and Benefits of Minimum Competency Testing</i>	355	
	<i>Perceived Drawbacks of Minimum Competency Testing</i>	356	
	<i>Responsibilities: State and Local Levels</i>	357	
	<i>Legal Principles of Minimum Competency Testing</i>	358	
	<i>Recent Criticisms of Minimum Competency Tests</i>	359	
	College Admissions Tests	360	
	SAT	361	
	ACT	361	



Comments on the Scholastic Aptitude Test	362
<i>Validity of the SAT</i>	362
<i>SAT and ACT Spin-Off Industries</i>	363
<i>Changes in the SAT and ACT</i>	364
Teacher Examinations	364
<i>National Teacher Examinations</i>	365
<i>Pre-Professional Skills Test</i>	366
<i>Questions That Need Answers</i>	366
Summary	368
Key Terms and Concepts	369
Review Items	369
Exercises	371
References	372

## **APPENDIX 1**

Code of Fair Testing Practices in Education	375
---	-----

## **APPENDIX 2**

Answers to Review Exercises	381
-----------------------------	-----

## **APPENDIX 3**

Solutions to Chapter Exercises	386
--------------------------------	-----

## **APPENDIX 4**

Table of the Normal Curve	396
---------------------------	-----

<b>GLOSSARY OF KEY TERMS AND CONCEPTS</b>	<b>399</b>
---	------------

<b>GLOSSARY OF STATISTICAL SYMBOLS AND FORMULAS</b>	<b>406</b>
---	------------

<b>AUTHOR INDEX</b>	<b>409</b>
---------------------	------------

<b>SUBJECT INDEX</b>	<b>411</b>
----------------------	------------