

Measurement and Assessment in Education

Cecil R. Reynolds

Texas A&M University

Ronald B. Livingston

University of Texas at Tyler

Victor Willson

Texas A&M University



Boston ■ New York ■ San Francisco
Mexico City ■ Montreal ■ Toronto ■ London ■ Madrid ■ Munich ■ Paris
Hong Kong ■ Singapore ■ Tokyo ■ Cape Town ■ Sydney

CONTENTS

Preface xiii

1 Introduction to Educational Assessment 1

The Language of Assessment 2

Tests, Measurement, and Assessment 3 / Types of Tests 4 / Types of Score Interpretations 8

Assumptions of Educational Assessment 9

Psychological and Educational Constructs Exist 9 / Psychological and Educational Constructs Can Be Measured 10 / Although We Can Measure Constructs, Our Measurement Is Not Perfect 10 / There Are Different Ways to Measure Any Given Construct 10 / All Assessment Procedures Have Strengths and Limitations 10 / Multiple Sources of Information Should Be Part of the Assessment Process 11 / Performance on Tests Can Be Generalized to Nontest Behaviors 11 / Assessment Can Provide Information That Helps Educators Make Better Educational Decisions 11 / Assessments Can Be Conducted in a Fair Manner 11 / Testing and Assessment Can Benefit Our Educational Institutions and Society as a Whole 12

Participants in the Assessment Process 13

People Who Develop Tests 13 / People Who Use Tests 14 / People Who Take Tests 14 / Other People Involved in the Assessment Process 15

Common Applications of Educational Assessments 15

Student Evaluations 15 / Instructional Decisions 16 / Selection, Placement, and Classification Decisions 16 / Policy Decisions 17 / Counseling and Guidance Decisions 18

What Teachers Need to Know about Assessment 18

Teachers Should Be Proficient in Selecting Professionally Developed Assessment Procedures Appropriate for Making Instructional Decisions 19 / Teachers Should Be Proficient in Developing Assessment Procedures Appropriate for Making Instructional Decisions 19 / Teachers Should Be Proficient in Administering, Scoring, and Interpreting Professionally Developed and Teacher-Made Assessment Procedures 20 / Teachers Should Be Proficient in Using Assessment Results When Making Educational Decisions 20 / Teachers Should Be Proficient in Developing Valid Grading Procedures That Incorporate Assessment Information 20 / Teachers Should Be Proficient in Communicating Assessment Results 20 / Teachers Should Be Proficient in Recognizing Unethical, Illegal, and Other Inappropriate Uses of Assessment Procedures or Information 21

Educational Assessment in the Twenty-First Century 21

Computerized Adaptive Testing (CAT) and Other Technological Advances 22 / "Authentic" or Complex-Performance Assessments 22 / Educational

Accountability and High-Stakes Assessment 24 / Trends in the Assessment
of Students with Disabilities 24

Summary 26

2 The Basic Mathematics of Measurement 31

The Role of Mathematics in Assessment 31

Scales of Measurement 32

What Is Measurement? 32 / Nominal Scales 33 / Ordinal Scales 33
Interval Scales 34 / Ratio Scales 34

The Description of Test Scores 36

Distributions 36 / Measures of Central Tendency 40 /
Measures of Variability 43

Correlation Coefficients 49

Scatterplots 50 / Correlation and Prediction 52 / Types of Correlation
Coefficients 52 / Correlation versus Causality 52

Summary 55

3 The Meaning of Test Scores 58

Norm-Referenced and Criterion-Referenced Score Interpretations 59

Norm-Referenced Interpretations 60 / Criterion-Referenced
Interpretations 76

Norm-Referenced, Criterion-Referenced, or Both? 79

Qualitative Description of Scores 81

Summary 82

4 Reliability for Teachers 85

Errors of Measurement 86

Sources of Measurement Error 87

Methods of Estimating Reliability 90

Test-Retest Reliability 92 / Alternate-Form Reliability 93 / Internal-
Consistency Reliability 93 / Inter-rater Reliability 97 / Reliability
of Composite Scores 99 / Selecting a Reliability Coefficient 99 /
Evaluating Reliability Coefficients 101 / How to Improve Reliability 103 /
Special Problems in Estimating Reliability 105

The Standard Error of Measurement 106

Evaluating the Standard Error of Measurement 108

	Reliability: Practical Strategies for Teachers	111
	Summary	114
5	Validity for Teachers	117
	Threats to Validity	118
	Reliability and Validity	119
	“Types of Validity” versus “Types of Validity Evidence”	120
	Types of Validity Evidence	123
	Evidence Based on Test Content 123 / Evidence Based on Relations to Other Variables 126 / Evidence Based on Internal Structure 133 / Evidence Based on Response Processes 134 / Evidence Based on Consequences of Testing 134 / Integrating Evidence of Validity 135	
	Validity: Practical Strategies for Teachers	137
	Summary	138
6	Item Analysis for Teachers	141
	Item Difficulty Index (or Item Difficulty Level)	142
	Special Assessment Situations and Item Difficulty	144
	Item Discrimination	144
	Discrimination Index 145 / Item-Total Correlation Coefficients 147 / Item Discrimination on Mastery Tests 149 / Item Analysis of Speed Tests 150	
	Distracter Analysis	151
	How Distracters Influence Item Difficulty and Discrimination	152
	Item Analysis: Practical Strategies for Teachers	153
	Using Item Analysis to Improve Items	155
	Item Analysis of Performance Assessments	157
	Qualitative Item Analysis	158
	Using Item Analysis to Improve Classroom Instruction	159
	Summary	161
7	The Initial Steps in Developing a Classroom Test	163
	Characteristics of Educational Objectives	165
	Scope	165

Taxonomy of Educational Objectives	166
Cognitive Domain	166 / Affective Domain 169 / Psychomotor Domain 170
Behavioral versus Nonbehavioral Educational Objectives	171
Writing Educational Objectives	172
Developing a Table of Specifications (or Test Blueprint)	173
Implementing the Table of Specifications and Developing an Assessment	175
Norm-Referenced versus Criterion-Referenced Score Interpretations	176 / Selecting Which Types of Items to Use 176 / Putting the Assessment Together 180
Preparing Your Students and Administering the Assessment	183
Summary	185
8 The Development and Use of Selected-Response Items	188
Multiple-Choice Items	189
Guidelines for Developing Multiple-Choice Items	190 / Strengths and Weaknesses of Multiple-Choice Items 199
True-False Items	204
Guidelines for Developing True-False Items	205 / Strengths and Weaknesses of True-False Items 206
Matching Items	208
Guidelines for Developing Matching Items	209 / Strengths and Weaknesses of Matching Items 211
Summary	213
9 The Development and Use of Constructed-Response Items	215
Oral Testing: The Oral Essay as a Precursor of Constructed-Response Items	216
Essay Items	217
Purposes of Essay Items	217 / Essay Items at Different Levels of Complexity 219 / Restricted-Response versus Extended-Response Essays 221 / Guidelines for Developing Essay Items 222 / Strengths and Weaknesses of Essay Items 223 / Guidelines for Scoring Essay Items 226
Short-Answer Items	230
Guidelines for Developing Short-Answer Items	232 / Strengths and Weaknesses of Short-Answer Items 234

A Final Note: Constructed-Response versus Selected-Response Items 235

Summary 236

10 Performance Assessments and Portfolios 238**What Are Performance Assessments? 239****Guidelines for Developing Effective Performance Assessments 245**

Selecting Appropriate Performance Tasks 245 / Developing Instructions 249

Developing Procedures for Evaluating Responses 249 / Implementing
Procedures to Minimize Errors in Rating 254**Strengths and Weaknesses of Performance Assessments 258****Portfolios 262**Guidelines for Developing Portfolio Assessments 262 / Strengths and
Weaknesses of Portfolio Assessments 264

Summary 266

**11 Assigning Grades on the Basis
of Classroom Assessments 270****Feedback and Evaluation 271**Formal and Informal Evaluation 274 / The Use of Formative Evaluation
in Summative Evaluation 274**Reporting Student Progress: Which Symbols to Use? 275****The Basis for Assigning Grades 277****Frame of Reference 278**Norm-Referenced Grading (Relative Grading) 278 / Criterion-Referenced
Grading (Absolute Grading) 280 / Achievement in Relation to Improvement or
Effort 281 / Achievement Relative to Ability 282 / Recommendation 282**Combining Grades into a Composite 283****Informing Students of Grading System 288****Parent Conferences 288**

Summary 289

**12 Standardized Achievement Tests in
the Era of High-Stakes Assessment 291****Group-Administered Achievement Tests 294**Commercially Developed Group Achievement Tests 295 / State-Developed
Achievement Tests 304 / Best Practices in Using Standardized Achievement
Tests in Schools 306

Individual Achievement Tests	315
Selecting an Achievement Battery	317
Summary	318

13 The Use of Aptitude Tests in the Schools 320

A Brief History of Intelligence Tests	323
The Use of Aptitude and Intelligence Tests in Schools	326
Aptitude–Achievement Discrepancies	327
Major Aptitude/Intelligence Tests	329
Group Aptitude/Intelligence Tests	329 / Individual Aptitude/Intelligence Tests 334 / Selecting Aptitude/Intelligence Tests 342
College Admission Tests	343
Summary	344

14 Assessment of Behavior and Personality 347

Assessing Behavior and Personality	349
Response Sets	349 / Assessment of Behavior and Personality in the Schools 350
Behavior Rating Scales	352
Behavior Assessment System for Children—Teacher Rating Scale and Parent Rating Scale (TRS and PRS)	353 / Conners Rating Scales—Revised (CRS-R) 357 / Child Behavior Checklist and Teacher Report Form (CBCL and TRF) 358
Self-Report Measures	360
Behavior Assessment System for Children—Self-Report of Personality (SRP)	360 / Youth Self-Report (YSR) 364
Projective Techniques	364
Projective Drawings	366 / Sentence Completion Tests 367 / Apperception Tests 367 / Inkblot Techniques 367
Summary	369

15 Assessment Accommodations for Students with Disabilities 371

Major Legislation That Impacts the Assessment of Students with Disabilities	373
--	------------

Individuals with Disabilities Education Act (IDEA)	373
IDEA Categories of Disabilities	375
Section 504	378
The Rationale for Assessment Accommodations	379
When Are Accommodations <i>Not</i> Appropriate or Necessary?	380
Strategies for Accommodations	380
Modifications of Presentation Format	381 / Modifications of Response
Format	381 / Modifications of Timing
Adaptive Devices and Supports	383 / Modification of Setting
Using Alternate Assessments	385 / Using Only a Portion of a Test
Using Alternate Assessments	385
Determining What Accommodations to Provide	385
Reporting Results of Modified Assessments	387
Summary	390
16 The Problem of Bias in Educational Assessment	395
What Do We Mean by Bias?	398
Past and Present Concerns: A Brief Look	399
The Controversy over Bias in Testing: Its Origin, What It Is, and What It Is Not	399
Cultural Bias and the Nature of Psychological Testing	405
Objections to the Use of Educational and Psychological Tests with Minority Students	406
Inappropriate Content	406 / Inappropriate Standardization Samples
Examiner and Language Bias	407 / Inequitable Social Consequences
Measurement of Different Constructs	407 / Differential Predictive
Validity	407 / Qualitatively Distinct Aptitude and Personality
Validity	407 / Qualitatively Distinct Aptitude and Personality
The Problem of Definition in Test Bias Research: Differential Validity	408
Cultural Loading, Cultural Bias, and Culture-Free Tests	408
Inappropriate Indicators of Bias: Mean Differences and Equivalent Distributions	409
Bias in Test Content	410
Bias in Other Internal Features of Tests	413
Bias in Prediction and in Relation to Variables External to the Test	415
Summary	420

17	Best Practices in Educational Assessment	422
	Guidelines for Developing Assessments	424
	Guidelines for Selecting Published Assessments	425
	Guidelines for Administering Assessments	429
	Guidelines for Scoring Assessments	432
	Guidelines for Interpreting, Using, and Communicating Assessment Results	434
	Responsibilities of Test Takers	435
	Summary	437
	APPENDIX A: Summary Statements of <i>The Student Evaluation Standards</i> (JCSEE, 2003)	441
	APPENDIX B: <i>Code of Professional Responsibilities in Educational Measurement</i> (NCME, 1995)	444
	APPENDIX C: <i>Code of Fair Testing Practices in Education</i> (JCTP, 1988)	452
	APPENDIX D: <i>Rights and Responsibilities of Test Takers: Guidelines and Expectations</i> (JCTP, 1998)	456
	APPENDIX E: <i>Standards for Teacher Competence in Educational Assessment of Students</i> (AFT, NCME, and NEA, 1990)	465
	APPENDIX F: Proportions of Area under the Normal Curve	471
	APPENDIX G: Answers to Practice Problems	475
	References	477
	Index	483