Designing and Using Tools for EDUCATIONAL ASSESSMENT

Madhabi Chatterji

CONTENTS

Preface ix

1 Assessing Educational Constructs: Basic Concepts 1

Overview 1

1.1 What Is Educational Assessment? 2

1.2 Elements of a Useful Assessment Procedure 6

1.3 Operational Definitions of Constructs 9

1.4 Constructs and Variables 11

1.5 Assessment, Measurement, and Evaluation 12

1.6 Role of Assessment in Education 13

Summary 19

2 Purposes for Educational Assessment 22

Overview 22

2.1 The Need for Clear Assessment Purposes 22

2.2 A Typology of Assessment Uses in Education 25

2.3 User Path 1: Assessment for Teaching and Learning 27

2.4 User Path 2: Assessment for Program Planning, Evaluation, and Policy Contexts 33

2.5 User Path 3: Assessment for Screening and Diagnosis of Exceptionalities 39

2.6 User Path 4: Assessment for Guidance and Counseling 43

2.7 User Path 5: Assessment for Admissions, Licensure, Scholarships, and Awards 46

2.8 User Path 6: Assessment for Educational Research and Development 48

2.9 Crossover across User Groups 48

2.10 Responsibilities for Appropriate Assessment Use 49

Summary 50

3 Quality of Assessment Results: Validity, Reliability, and Utility 53

Overview 53

3.1 Validity 54

3.2 Validation and Types of Validity Evidence 58

3.3 Validation: When Should We Do It and How Far Should We Go? 65

3.4 Reliability 67

3.5 Utility 70

3.6 Prioritizing among Validity, Reliability, and Utility 73

Summary 74

4 Types of Assessment Tools 78

Overview 78

4.1 Traditional, Alternative, Authentic, and Performance Assessments 79

4.2 Other Ways of Classifying Assessments 83

4.3 Types of Assessments Based on Mode of Response 87

4.4 Advantages and Disadvantages of Different Assessment Methods 95

Summary 100

5 A Process Model for Designing, Selecting, and Validating Assessment Tools 103

Overview 103

5.1 Need for a General Process Model 104

5.2 Components of a Process Model for Assessment Design/Selection and Validation 105

5.3 A Case Study in Using the Process Model 111

5.4 The Importance of Following a Systematic Process 117

Summary 119

6 Specifying the Construct Domain 121

Overview 121

6.1 Specifying the Domain for Constructs in User Path 1 122

6.2 Taxonomies of Learning Outcomes 137

6.3 Specifying the Domain for Constructs in User Paths 2–6 142

Summary 149

7 Designing or Selecting Written Structured-Response Assessment Tools 152

Overview 152

7.1 Why Use Written Structured-Response Assessment Tools? 153

7.2 The Process Model Applied to W-SR Assessment Design or Selection 155

7.3 Developing Assessment Specifications for W-SR Tools 158

7.4 Guidelines for Item Construction for Different Types of W-SR Items 165

7.5 Complex Interpretive W-SR Exercises 187

7.6 The Last Word on Clues 190

7.7 Choosing the Best W-SR Item Format 190

7.8 W-SR Test Assembly 191

7.9 Content Validation 193

7.10 Bias During W-SR Assessment Design 195

Summary 196

8 Designing or Selecting Performance Assessments 200

Overview 200

8.1 Why Use Performance Assessments? 201

8.2 Justifying Performance Assessment Methods We Choose 202

8.3 Applying the Process Model with Performance Assessments 206

8.4 Designing Different Types of Performance Assessments 215
8.5 Scoring Rubrics and How to Develop Them 234
8.6 Specifications for Performance Assessments 239
8.7 Assembling and Content-Validating Performance Assessments 243
8.8 Using Rubrics: Sources of Random and Systematic Error 244
Summary 247

9 Designing or Selecting Affective, Social-Emotional, Personality, and Behavioral Assessments 250

Overview 250
9.1 Applying the Process Model in User Paths 2–6 251
9.2 The Nature of Constructs in Paths 2–6 262
9.3 Designing Self-Report Instruments 263
9.4 *Designing Structured Observation Forms* 277
9.5 Naturalistic and Anecdotal Observations 279
9.6 Selecting Assessment Tools Using Specifications 282
9.7 Classical Examples of Instrument Design in Paths 2–6 284
Summary 286

10 Analyzing Data from Assessments 289

Overview 289
10.1 Scales of Measurement 290
10.2 Continuous and Discontinuous Variables 294
10.3 Organizing Data 295
10.4 Measures of Central Tendency 301
10.5 Measures of Variability 309
10.6 Graphical Displays of Distributions 312
10.7 Normal Distribution and Its Applications 318
10.8 Skewness and Kurtosis in Distributions 320

10.9 Measures of Relative Position 322

10.10 Correlation Coefficients and Their Applications 328

Summary 338

11 Decision-Making Applications in Different User Paths 342

Overview 342

11.1 Setting Standards 343

11.2 Report Card Marking 353

11.3 Domain-Referenced Mastery Analysis 364

11.4 Mapping Long-Term Trends on Measured Constructs 371

11.5 Using Assessment Results for Planning, Determining Needs, or Evaluating Programs/Services 376

Summary 381

12 Quantitative Item Analysis 383

Overview 383

12.1 Purposes for Item Analysis 384

12.2 Item Analysis Indices 384

12.3 Differences between NRT and CRT Item Analysis 387

12.4 Application of Item Analysis for NRTs 388

12.5 Application of Item Analysis for CRTs 394

12.6 Item Descriptive Statistics 400

12.7 Limitations of Item Analysis Studies 403

Summary 404

13 Quantitative Evaluation of Validity and Reliability 407

Overview 407

13.1 Empirical Validation Methods 408

13.2 Empirical Estimation of Reliability 427

13.3 Reliability in Criterion-Referenced Measurements 441

Summary 443

14 Selecting and Using Standardized Assessment Tools 445

Overview 445

14.1 Distinguishing Characteristics of Standardized Tools 446

14.2 Standardized Tools in Use in Education 448

14.3 Norms, Norm-Referenced Scores, and Score Profiles 464

14.4 Evaluating Standards-Based Assessments 477

14.5 Resources for Finding Published Assessment Tools 480

Summary 484

Bibliography 486

Index 493