

CONTENTS

List of Figures, Tables, and Sidebars	xxiii
Introduction: A Little Knowledge Is Dangerous	1
Why Test?	1
Why Read This Book?	2
A Confusing State of Affairs	3
Misleading Familiarity	3
Inaccessible Technology	4
Procedural Confusion	4
Testing and Kirkpatrick's Levels of Evaluation	5
Certification in the Corporate World	7
Corporate Testing Enters the New Millennium	10
What Is to Come. . .	11
PART I: BACKGROUND: THE FUNDAMENTALS	13
1 Test Theory	15
What Is Testing?	15
What Does a Test Score Mean?	17
Reliability and Validity: A Primer	18
Reliability	18
Equivalence Reliability	19
Test-Retest Reliability	19
Inter-Rater Reliability	19
Validity	20
Face Validity	23
Context Validity	23
Concurrent Validity	23
Predictive Validity	24
Concluding Comment	24

2	Types of Tests	25
	Criterion-Referenced Versus Norm-Referenced Tests	25
	Frequency Distributions	25
	Criterion-Referenced Test Interpretation	28
	Six Purposes for Tests in Training Settings	30
	Three Methods of Test Construction (One of Which You Should Never Use)	32
	Topic-Based Test Construction	32
	Statistically Based Test Construction	33
	Objectives-Based Test Construction	34
	PART II: OVERVIEW: THE CRTD MODEL AND PROCESS	37
3	The CRTD Model and Process	39
	Relationship to the Instructional Design Process	39
	The CRTD Process	43
	Plan Documentation	44
	Analyze Job Content	44
	Establish Content Validity of Objectives	46
	Create Items	46
	Create Cognitive Items	46
	Create Rating Instruments	47
	Establish Content Validity of Items and Instruments	47
	Conduct Initial Test Pilot	47
	Perform Item Analysis	48
	Difficulty Index	48
	Distractor Pattern	48
	Point-Biserial	48
	Create Parallel Forms or Item Banks	49
	Establish Cut-Off Scores	49
	Informed Judgment	50
	Angoff	50

Contrasting Groups	50
Determine Reliability	50
Determine Reliability of Cognitive Tests	50
Equivalence Reliability	51
Test-Retest Reliability	51
Determine Reliability of Performance Tests	52
Report Scores	52
Summary	53

PART III: THE CRTD PROCESS: PLANNING AND

CREATING THE TEST 55

4	Plan Documentation	57
	Why Document?	57
	What to Document	63
	The Documentation	64
5	Analyze Job Content	75
	Job Analysis	75
	Job Analysis Models	77
	Summary of the Job Analysis Process	78
	DACUM	79
	Hierarchies	87
	Hierarchical Analysis of Tasks	87
	Matching the Hierarchy to the Type of Test	88
	Prerequisite Test	89
	Entry Test	89
	Diagnostic Test	89
	Posttest	89
	Equivalency Test	90
	Certification Test	90
	Using Learning Task Analysis to	91
	Validate a Hierarchy	
	Bloom's Original Taxonomy	91
	Knowledge Level	92

	Comprehension Level	93
	Application Level	93
	Analysis Level	93
	Synthesis Level	93
	Evaluation Level	94
	Using Bloom’s Original Taxonomy to Validate a Hierarchy	94
	Bloom’s Revised Taxonomy	95
	Gagné’s Learned Capabilities	96
	Intellectual Skills	96
	Cognitive Strategies	97
	Verbal Information	97
	Motor Skill	97
	Attitudes	97
	Using Gagné’s Intellectual Skills to Validate a Hierarchy	97
	Merrill’s Component Design Theory	98
	The Task Dimension	99
	Types of Learning	99
	Using Merrill’s Component Design Theory to Validate a Hierarchy	99
	Data-Based Methods for Hierarchy Validation	100
	Who Killed Cock Robin?	102
6	Content Validity of Objectives	105
	Overview of the Process	105
	The Role of Objectives in Item Writing	106
	Characteristics of Good Objectives	107
	Behavior Component	107
	Conditions Component	108
	Standards Component	108
	A Word from the Legal Department About Objectives	109
	The Certification Suite	109

	Certification Levels in the Suite	110
	Level A—Realworld	110
	Level B—High-Fidelity Simulation	111
	Level C—Scenarios	111
	Quasi-Certification	112
	Level D—Memorization	112
	Level E—Attendance	112
	Level F—Affiliation	113
	How to Use the Certification Suite	113
	Finding a Common Understanding	113
	Making a Professional Decision	114
	The correct level to match the job	114
	The operationally correct level	114
	The consequences of lower fidelity	115
	Converting Job-Task Statements to Objectives	116
	In Conclusion	119
7	Create Cognitive Items	121
	What Are Cognitive Items?	121
	Classification Schemes for Objectives	122
	Bloom’s Cognitive Classifications	123
	Types of Test Items	129
	Newer Computer-Based Item Types	129
	The Six Most Common Item Types	130
	True/False Items	131
	Matching Items	132
	Multiple-Choice Items	132
	Fill-In Items	147
	Short Answer Items	147
	Essay Items	148
	The Key to Writing Items That Match Jobs	149
	The Single Most Useful Improvement You	149
	Can Make in Test Development	
	Intentional Versus Extensional Items	150
	Show Versus Tell	152

	The Certification Suite	155
	Guidelines for Writing Test Items	158
	Guidelines for Writing the Most Common Item Types	159
	How Many Items Should Be on a Test?	166
	Test Reliability and Test Length	166
	Criticality of Decisions and Test Length	167
	Resources and Test Length	168
	Domain Size of Objectives and Test Length	168
	Homogeneity of Objectives and Test Length	169
	Research on Test Length	170
	Summary of Determinants of Test Length	170
	A Cookbook for the SME	172
	Deciding Among Scoring Systems	174
	Hand Scoring	175
	Optical Scanning	175
	Computer-Based Testing	176
	Computerized Adaptive Testing	180
8	Create Rating Instruments	183
	What Are Performance Tests?	183
	Product Versus Process in Performance Testing	187
	Four Types of Rating Scales for Use in Performance Tests (Two of Which You Should Never Use)	187
	Numerical Scales	188
	Descriptive Scales	188
	Behaviorally Anchored Rating Scales	188
	Checklists	190
	Open Skill Testing	192
9	Establish Content Validity of Items and Instruments	195
	The Process	195

	Establishing Content Validity—The Single Most Important Step	196
	Face Validity	196
	Content Validity	197
	Two Other Types of Validity	202
	Concurrent Validity	202
	Predictive Validity	208
	Summary Comment About Validity	209
10	Initial Test Pilot	211
	Why Pilot a Test?	211
	Six Steps in the Pilot Process	212
	Determine the Sample	212
	Orient the Participants	213
	Give the Test	214
	Analyze the Test	214
	Interview the Test-Takers	215
	Synthesize the Results	216
	Preparing to Collect Pilot Test Data	217
	Before You Administer the Test	217
	Sequencing Test Items	217
	Test Directions	218
	Test Readability Levels	219
	Lexile Measure	220
	Formatting the Test	220
	Setting Time Limits—Power, Speed, and Organizational Culture	221
	When You Administer the Test	222
	Physical Factors	222
	Psychological Factors	222
	Giving and Monitoring the Test	223
	Special Considerations for Performance Tests	225
	Honesty and Integrity in Testing	231
	Security During the Training-Testing Sequence	234

	Organization-Wide Policies Regarding Test Security	236
11	Statistical Pilot	241
	Standard Deviation and Test Distributions	241
	The Meaning of Standard Deviation	241
	The Five Most Common Test Distributions	244
	Problems with Standard Deviations and Mastery Distributions	247
	Item Statistics and Item Analysis	248
	Item Statistics	248
	Difficulty Index	248
	P-Value	249
	Distractor Pattern	249
	Point-Biserial Correlation	250
	Item Analysis for Criterion-Referenced Tests	251
	The Upper-Lower Index	253
	Phi	255
	Choosing Item Statistics and Item Analysis Techniques	255
	Garbage In-Garbage Out	257
12	Parallel Forms	259
	Paper-and-Pencil Tests	260
	Computerized Item Banks	262
	Reusable Learning Objects	264
13	Cut-Off Scores	265
	Determining the Standard for Mastery	265
	The Outcomes of a Criterion-Referenced Test	266
	The Necessity of Human Judgment in Setting a Cut-Off Score	267
	Consequences of Misclassification	267
	Stakeholders	268
	Revisability	268
	Performance Data	268

	Three Procedures for Setting the Cut-Off Score	269
	The Issue of Substitutability	269
	Informed Judgment	270
	A Conjectural Approach, the Angoff Method	272
	Contrasting Groups Method	278
	Borderline Decisions	282
	The Meaning of Standard Error of Measurement	282
	Reducing Misclassification Errors at the Borderline	284
	Problems with Correction-for-Guessing	285
	The Problem of the Saltatory Cut-Off Score	287
14	Reliability of Cognitive Tests	289
	The Concepts of Reliability, Validity, and Correlation	289
	Correlation	290
	Types of Reliability	293
	Single-Test-Administration Reliability Techniques	294
	Internal Consistency	294
	Squared-Error Loss	296
	Threshold-Loss	296
	Calculating Reliability for Single-Test Administration Techniques	297
	Livingston's Coefficient kappa (κ^2)	297
	The Index S_c	297
	Outcomes of Using the Single-Test- Administration Reliability Techniques	298
	Two-Test-Administration Reliability Techniques	299
	Equivalence Reliability	299
	Test-Retest Reliability	300
	Calculating Reliability for Two-Test Administration Techniques	301

	The Phi Coefficient	302
	Description of Phi	302
	Calculating Phi	302
	How High Should Phi Be?	304
	The Agreement Coefficient	306
	Description of the Agreement Coefficient	306
	Calculating the Agreement Coefficient	307
	How High Should the Agreement Coefficient Be?	308
	The Kappa Coefficient	308
	Description of Kappa	308
	Calculating the Kappa Coefficient	309
	How High Should the Kappa Coefficient Be?	311
	Comparison of ϕ , ρ_0 , and κ	313
	The Logistics of Establishing Test Reliability	314
	Choosing Items	314
	Sample Test-Takers	315
	Testing Conditions	316
	Recommendations for Choosing a Reliability Technique	316
	Summary Comments	317
15	Reliability of Performance Tests	319
	Reliability and Validity of Performance Tests	319
	Types of Rating Errors	320
	Error of Standards	320
	Halo Error	321
	Logic Error	321
	Similarity Error	321
	Central Tendency Error	321
	Leniency Error	322
	Inter-Rater Reliability	322
	Calculating and Interpreting Kappa (κ)	323
	Calculating and Interpreting Phi (ϕ)	335

	Repeated Performance and Consecutive Success	344
	Procedures for Training Raters	347
	What If a Rater Passes Everyone Regardless of Performance?	349
	What Should You Do?	352
	What If You Get a High Percentage of Agreement Among Raters But a Negative Phi Coefficient?	353
16	Report Scores	357
	CRT Versus NRT Reporting	358
	Summing Subscores	358
	What Should You Report to a Manager?	361
	Is There a Legal Reason to Archive the Tests?	362
	A Final Thought About Testing and Teaching	362
	PART IV: LEGAL ISSUES IN CRITERION-REFERENCED TESTING	365
17	Criterion-Referenced Testing and Employment Selection Laws	367
	What Do We Mean by Employment Selection Laws?	368
	Who May Bring a Claim?	368
	A Short History of the <i>Uniform Guidelines on Employee Selection Procedures</i>	370
	Purpose and Scope	371
	Legal Challenges to Testing and the <i>Uniform Guidelines</i>	373
	Reasonable Reconsideration	376
	In Conclusion	376
	Balancing CRTs with Employment Discrimination Laws	376
	Watch Out for Blanket Exclusions in the Name of Business Necessity	378
	Adverse Impact, the Bottom Line, and Affirmative Action	380

Adverse Impact	380
The Bottom Line	383
Affirmative Action	385
Record-Keeping of Adverse Impact and Job-Relatedness of Tests	387
Accommodating Test-Takers with Special Needs	387
Testing, Assessment, and Evaluation for Disabled Candidates	390
Test Validation Criteria: General Guidelines	394
Test Validation: A Step-by-Step Guide	397
1. Obtain Professional Guidance	397
2. Select a Legally Acceptable Validation Strategy for Your Particular Test	397
3. Understand and Employ Standards for Content-Valid Tests	398
4. Evaluate the Overall Test Circumstances to Assure Equality of Opportunity	399
Keys to Maintaining Effective and Legally Defensible Documentation	400
Why Document?	400
What Is Documentation?	401
Why Is Documentation an Ally in Defending Against Claims?	401
How Is Documentation Used?	402
Compliance Documentation	402
Documentation to Avoid Regulatory Penalties or Lawsuits	404
Use of Documentation in Court	404
Documentation to Refresh Memory	404
Documentation to Attack Credibility	404
Disclosure and Production of Documentation	405
Pay Attention to Document Retention Policies and Protocols	407
Use Effective Word Management in Your	409

Documentation	
Use Objective Terms to Describe Events and Compliance	412
Avoid Inflammatory and Off-the-Cuff Commentary	412
Develop and Enforce Effective Document Retention Policies	413
Make Sure Your Documentation Is Complete	414
Make Sure Your Documentation Is Capable of “Authentication”	415
In Conclusion	415
Is Your Criterion-Referenced Testing Legally Defensible? A Checklist	416
A Final Thought	419
Epilogue: CRTD as Organizational Transformation	421
References	425
Index	433
About the Authors	453

