

CONTENTS

<i>List of figures</i>	xvi
<i>List of tables</i>	xx
<i>Preface</i>	xxii
<i>Aims of the Book</i>	xxii
<i>Audiences for the Book</i>	xxiii
<i>Structure of the Book</i>	xxiii
<i>Learning Tools</i>	xxiv
<i>Using the Book to Teach a Course</i>	xxv
<i>Help for Old Friends</i>	xxvi
<i>Acknowledgements</i>	xxviii

PART I

A Constructive Approach to Measurement 1

1	The BEAR Assessment System: Overview of the “Four Building Blocks” Approach	3
1.1	What Is “Measurement”? 3	
1.1.1	Construct Modeling 6	
1.2	The BEAR Assessment System 7	
1.3	The Construct Map 8	
1.3.1	Example 1: The MoV Construct in the Data Modeling Assessments 11	

1.4	<i>The Items Design</i>	15
1.4.1	<i>Example 1: MoV Items</i>	16
1.4.2	<i>The Relationship between the Construct and the Responses</i>	18
1.5	<i>The Outcome Space</i>	19
1.5.1	<i>Example 1: The MoV Outcome Space</i>	21
1.6	<i>The Wright Map</i>	25
1.6.1	<i>Example 1: The MoV Wright Map</i>	26
1.6.2	<i>Return to the Discussion of Causation and Inference</i>	31
1.7	<i>Reporting the Results to the Measurer and Other Users</i>	32
1.8	<i>Using the Four Building Blocks to Develop an Instrument</i>	34
1.9	<i>Resources</i>	37
1.10	<i>Exercises and Activities</i>	37

PART II

The Four Building Blocks

39

2 Construct Maps

41

2.1	<i>The Construct Map</i>	41
2.2	<i>Examples of Construct Maps</i>	45
2.2.1	<i>Example 1: The Models of Variability (MoV) Construct in the Data Modeling Curriculum</i>	46
2.2.2	<i>Example 2: A Social and Emotional Learning Example (RIS: The Researcher Identity Scale)</i>	47
2.2.3	<i>Example 3: An Attitude Example (GEB: General Ecological Behavior)</i>	48
2.2.4	<i>Example 4: A 21st Century Skills Example (LPS Argumentation)</i>	51
2.2.5	<i>Example 5: The Six Constructs in the Data Modeling Curriculum</i>	53
2.2.6	<i>Example 6: A Process Measurement Example— Collaborative Problem-Solving (CPS)</i>	56
2.2.7	<i>Example 7: A Health Assessment Example (PF-10: Physical Functioning 10)</i>	58

2.2.8	<i>Example 8: An Interview Example (CUE: Conceptual Underpinnings of Evolution)</i>	60
2.2.9	<i>Example 9: An Observational Instrument—Early Childhood (DRDP)</i>	62
2.2.10	<i>Example 10: The Issues Evidence and You (IEY) Science Assessment</i>	64
2.3	<i>Using Construct Mapping to Help Develop an Instrument</i>	65
2.4	<i>Examples of Other Construct Structures</i>	67
2.5	<i>Resources</i>	69
2.6	<i>Exercises and Activities</i>	69
3	The Items Design	71
3.1	<i>The Idea of an Item</i>	71
3.2	<i>The Facets of the Items Design</i>	74
3.2.1	<i>The Construct Facet</i>	75
3.2.2	<i>The Secondary Design Facets</i>	78
3.3	<i>Different Types of Item Responses</i>	81
3.3.1	<i>Participant Observation</i>	83
3.3.2	<i>Specifying (Just) the Topics</i>	83
3.3.3	<i>Constructed Response Items</i>	84
3.3.4	<i>Selected Response Items</i>	85
3.3.5	<i>Steps in Item Development</i>	88
3.4	<i>Building-in Fairness through Design</i>	90
3.4.1	<i>What Do We Mean by Fairness Here?</i>	90
3.4.2	<i>Universal Design</i>	93
3.5	<i>Resources</i>	94
3.6	<i>Exercises and Activities</i>	94
4	The Outcome Space	96
4.1	<i>The Qualities of an Outcome Space</i>	96
4.1.1	<i>Well-defined Categories</i>	98
4.1.2	<i>Research-based Categories</i>	99
4.1.3	<i>Context-specific Categories</i>	100
4.1.4	<i>Finite and Exhaustive Categories</i>	101
4.1.5	<i>Ordered Categories</i>	102
4.2	<i>Scoring the Outcome Space (the Scoring Guide)</i>	103
4.3	<i>General Approaches to Constructing an Outcome Space</i>	104

6.3	<i>Resources</i>	190
6.4	<i>Exercises and Activities</i>	191
	<i>Textbox 6.1 The Partial Credit Model</i>	171
	<i>Textbox 6.2 Calculating the Thurstonian Thresholds</i>	174
7	Trustworthiness, Precision, and Reliability	193
7.1	<i>Trustworthiness in Measurement</i>	193
7.2	<i>Measurement Error: Precision</i>	195
7.3	<i>Summaries of Measurement Error</i>	203
	7.3.1 <i>Internal Consistency Coefficients</i>	204
	7.3.2 <i>Test–Retest Coefficients</i>	206
	7.3.3 <i>Alternate Forms Coefficients</i>	207
	7.3.4 <i>Other Reliability Coefficients and Indexes</i>	207
7.4	<i>Inter-rater Consistency</i>	209
7.5	<i>Resources</i>	212
7.6	<i>Exercises and Activities</i>	212
8	Trustworthiness, Validity, and Fairness	215
8.1	<i>Trustworthiness, Continued</i>	215
	8.1.1 <i>Crafting a Full Validity Argument</i>	216
8.2	<i>Evidence Based on Instrument Content</i>	217
	8.2.1 <i>Instrument Content Evidence for Example 2, the Researcher Identity Scale-G</i>	219
8.3	<i>Evidence Based on Response Processes</i>	222
	8.3.1 <i>Response Process Evidence Related to Example 8—The DRDP</i>	223
8.4	<i>Evidence Based on Internal Structure</i>	224
	8.4.1 <i>Evidence of Internal Structure at the Instrument Level: Dimensionality</i>	225
	8.4.2 <i>Dimensionality Evidence for Example 2: The Researcher Identity Scale-G</i>	227
	8.4.3 <i>Evidence of Internal Structure at the Instrument Level: The Wright Map</i>	229
	8.4.4 <i>Wright Map Evidence from Example 2: The Researcher Identity Scale-G</i>	230
	8.4.5 <i>Evidence of Internal Structure at the Item Level</i>	230

- 8.4.6 *Item-level Evidence of Internal Structure for the PF-10 Instrument* 232
- 8.5 *Evidence Regarding Relations to Other Variables* 233
 - 8.5.1 *“Other Variables” Evidence from Two Examples* 235
- 8.6 *Evidence Based on the Consequences of Using the Instrument* 237
- 8.7 *Evidence Related to Fairness* 238
 - 8.7.1 *Differential Item Functioning (DIF)* 239
 - 8.7.2 *DIF Evidence for the RIS-G* 240
- 8.8 *Resources* 242
- 8.9 *Exercises and Activities* 242

PART IV

A Beginning Rather than a Conclusion 245

9 *Building on the Building Blocks* 247

- 9.1 *Choosing the Statistical Model* 247
 - 9.1.1 *Interpretation of Thurstone’s Requirement in Terms of the Construct Map* 252
- 9.2 *Comparing Overall Model Fit* 257
- 9.3 *Beyond the Lone Construct Map: Multidimensionality* 259
- 9.4 *Resources* 267
- 9.5 *Exercises and Activities* 269
 - Textbox 9.1 Showing that Equation 9.5 Holds for the Rasch Model* 250
 - Textbox 9.2 Statistical Formulation of the Multidimensional Partial Credit Model* 260

10 *Beyond the Building Blocks* 271

- 10.1 *Beyond the Construct Map: Learning Progressions* 271
- 10.2 *Beyond the Items Design and the Outcome Space: Process Measurement* 275
- 10.3 *Beyond the Statistical Model: Considering a More Complex Scientific Model* 282
- 10.4 *Other Measurement Frameworks: Principled Assessment Designs* 291

10.4.1	<i>Example: Evidence-centered Design</i>	292
10.4.2	<i>Going "Outside the Triangle"</i>	294
10.5	<i>A Beginning Rather Than a Conclusion</i>	298
10.5.1	<i>Further Reading about the History of Measurement in the Social Sciences</i>	298
10.5.2	<i>Further Reading about Alternative Approaches</i>	299
10.5.3	<i>Further Reading about the Philosophy of Measurement</i>	300
10.6	<i>Exercises and Activities</i>	301
Appendix A	The Examples Archive	302
Appendix B	Computerized Design, Development, Delivery, Scoring and Reporting—BASS	304
Appendix C	The BEAR Assessment System (BAS): Papers about its Uses and Applications	306
Appendix D	Models of Variation Materials	316
Appendix E	The General Ecological Behavior Items	319
Appendix F	The Item Panel	321
Appendix G	Matching Likert and Guttman Items in the RIS Example	324
Appendix H	Sample Script for a Think-aloud Investigation	328
Appendix I	The Item Pilot Investigation	331
Appendix J	Results from the PF-10 Analyses	333
	<i>References</i>	336
	<i>Index</i>	353