

CONTENTS

Acknowledgments	xv
A Foreword (in Berkian Style) by Mike Theall	xix
Introduction	1
1 TOP 13 SOURCES OF EVIDENCE OF TEACHING EFFECTIVENESS	9
<i>A Few Ground Rules</i>	10
<i>Teaching Effectiveness: Defining the Construct</i>	11
National Standards	12
Beyond Student Ratings	13
A Unified Conceptualization	13
<i>Thirteen Sources of Evidence</i>	14
Student Ratings	15
Peer Ratings	19
External Expert Ratings	22
Self-Ratings	23
Videos	24
Student Interviews	25
Exit and Alumni Ratings	27
Employer Ratings	28
Administrator Ratings	29
Teaching Scholarship	30
Teaching Awards	31
Learning Outcome Measures	32

Teaching Portfolio	34
BONUS: 360° Multisource Assessment	37
<i>Berk's Top Picks</i>	44
Formative Decisions	45
Summative Decisions	45
Program Decisions	45
<i>Decision Time</i>	45
2 CREATING THE RATING SCALE STRUCTURE	47
<i>Overview of the Scale Construction Process</i>	48
<i>Specifying the Purpose of the Scale</i>	48
<i>Delimiting What Is to Be Measured</i>	50
Focus Groups	50
Interviews	50
Research Evidence	51
<i>Determining How to Measure the "What"</i>	54
Existing Scales	55
Item Banks	55
Commercially Published Student Rating Scales	56
<i>Universe of Items</i>	58
<i>Structure of Rating Scale Items</i>	60
Structured Items	60
Unstructured Items	62
3 GENERATING THE STATEMENTS	65
<i>Preliminary Decisions</i>	66
Domain Specifications	66
Number of Statements	66
<i>Rules for Writing Statements</i>	66
1. The statement should be <i>clear and direct</i> .	69
2. The statement should be <i>brief and concise</i> .	69
3. The statement should contain only <i>one complete behavior, thought, or concept</i> .	71
4. The statement should be a <i>simple sentence</i> .	72
5. The statement should be at the <i>appropriate reading level</i> .	72
6. The statement should be <i>grammatically correct</i> .	73
7. The statement should be <i>worded strongly</i> .	73
8. The statement should be <i>congruent with the behavior</i> it is intended to measure.	74

- | | |
|---|----|
| 9. The statement should accurately measure a <i>positive or negative behavior</i> . | 74 |
| 10. The statement should be <i>applicable to all respondents</i> . | 75 |
| 11. The respondents should be in the <i>best position to respond</i> to the statement. | 76 |
| 12. The statement should be <i>interpretable in only one way</i> . | 78 |
| 13. The statement should NOT contain a <i>double negative</i> . | 78 |
| 14. The statement should NOT contain <i>universal or absolute terms</i> . | 79 |
| 15. The statement should NOT contain <i>nonabsolute, warm-and-fuzzy terms</i> . | 79 |
| 16. The statement should NOT contain <i>value-laden or inflammatory words</i> . | 80 |
| 17. The statement should NOT contain <i>words, phrases, or abbreviations</i> that would be <i>unfamiliar to all respondents</i> . | 81 |
| 18. The statement should NOT tap a <i>behavior appearing in any other statement</i> . | 81 |
| 19. The statement should NOT be <i>factual</i> or capable of being interpreted as factual. | 82 |
| 20. The statement should NOT be <i>endorsed</i> or given one answer <i>by almost all respondents or by almost none</i> . | 83 |

4 SELECTING THE ANCHORS 85

- | | |
|---|-----|
| <i>Types of Anchors</i> | 86 |
| Intensity Anchors | 86 |
| Evaluation Anchors | 87 |
| Frequency Anchors | 89 |
| Quantity Anchors | 90 |
| Comparison Anchors | 91 |
| <i>Rules for Selecting Anchors</i> | 94 |
| 1. The anchors should be <i>consistent with the purpose</i> of the rating scale. | 94 |
| 2. The anchors should <i>match the statements, phrases, or word topics</i> . | 96 |
| 3. The anchors should be <i>logically appropriate</i> with each statement. | 98 |
| 4. The anchors should be <i>grammatically consistent</i> with each question. | 99 |
| 5. The anchors should provide the most <i>accurate and concrete</i> responses possible. | 100 |

6.	The anchors should elicit a <i>range of responses</i> .	101
7.	The anchors on <i>bipolar scales</i> should be <i>balanced, not biased</i> .	101
8.	The anchors on <i>unipolar scales</i> should be <i>graduated appropriately</i> .	102
5	REFINING THE ITEM STRUCTURE	105
	<i>Preparing for Structural Changes</i>	106
	<i>Issues in Scale Construction</i>	108
1.	What rating scale format is best?	108
2.	How many anchor points should be on the scale?	109
3.	Should there be a designated midpoint position, such as "Neutral," "Uncertain," or "Undecided," on the scale?	110
4.	How many anchors should be specified on the scale?	111
5.	Should numbers be placed on the anchor scale?	112
6.	Should a "Not Applicable" (NA) or "Not Observed" (NO) option be provided?	113
7.	How can response set biases be minimized?	115
6	ASSEMBLING THE SCALE FOR ADMINISTRATION	121
	<i>Assembling the Scale</i>	122
	Identification Information	122
	Purpose	123
	Directions	124
	Structured Items	125
	Unstructured Items	129
	<i>Scale Administration</i>	130
	Paper-Based Administration	131
	Online Administration	133
	Comparability of Paper-Based and Online Ratings	137
	Conclusions	139
7	FIELD TESTING AND ITEM ANALYSES	141
	<i>Preparing the Draft Scale for a Test Spin</i>	142
	<i>Field Test Procedures</i>	143
	Mini-Field Test	143
	Monster-Field Test	147
	<i>Item Analyses</i>	148
	Stage 1: Item Descriptive Statistics	148
	Stage 2: Interitem and Item-Scale Correlations	152
	Stage 3: Factor Analysis	155

8	COLLECTING EVIDENCE OF VALIDITY AND RELIABILITY	161
	<i>Validity Evidence</i>	162
	Evidence Based on Job Content Domain	164
	Evidence Based on Response Processes	168
	Evidence Based on Internal Scale Structure	169
	Evidence Related to Other Measures of Teaching Effectiveness	171
	Evidence Based on the Consequences of Ratings	172
	<i>Reliability Evidence</i>	174
	Classical Reliability Theory	174
	Summated Rating Scale Theory	175
	Methods for Estimating Reliability	176
	<i>Epilogue</i>	182
9	REPORTING AND INTERPRETING SCALE RESULTS	185
	<i>Generic Levels of Score Reporting</i>	186
	Item Anchor	186
	Item	187
	Subscale	190
	Total Scale	191
	Department/Program Norms	192
	Subject Matter/Program-Level State, Regional, and National Norms	193
	<i>Criterion-Referenced versus Norm-Referenced Score Interpretations</i>	194
	Score Range	194
	Criterion-Referenced Interpretations	195
	Norm-Referenced Interpretations	196
	<i>Formative, Summative, and Program Decisions</i>	197
	Formative Decisions	198
	Summative Decisions	204
	Program Decisions	212
	<i>Conclusions</i>	213
	References	215
	Appendices	241
	A. Sample "Home-Grown" Rating Scales	241
	B. Sample 360° Assessment Rating Scales	257
	C. Sample Reporting Formats	273
	D. Commercially Published Student Rating Scale Systems	277
	Index	281