

W. JAMES POPHAM

CLASSROOM ASSESSMENT

What Teachers
Need to Know

EIGHTH EDITION



e i g h t h e d i t i o n

Classroom Assessment

What Teachers Need to Know

W. James Popham

Professor Emeritus,
University of California, Los Angeles

PEARSON

Boston Columbus Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Vice President and Editorial Director:
Jeffery W. Johnston
Vice President and Publisher:
Kevin M. Davis
Editorial Assistant: Marisia Styles
Executive Field Marketing Manager: Krista
Clark
Senior Product Marketing Manager:
Christopher Barry
Project Manager: Pamela Bennett
Program Program Manager: Janelle Criner

Operations Specialist: Carol Melville
Text Designer: Cenveo
Cover Design Director: Diane Lorenzo
Cover Art: Getty Images, ©KidStock
Media Project Manager: Lauren Carlson
Full-Service Project Management: Cenveo
Composition: Cenveo
Printer/Binder: RR Donnelley/Harrisonburg
South
Cover Printer: Phoenix Color/Hagarstown
Text Font: 10.25/12 Sabon LT Std

Copyright © 2017, 2014, 2011 by Pearson Education, Inc. or its affiliates. All Rights Reserved. Printed in the United States of America. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions department, please visit www.pearsoned.com/permissions/.

Acknowledgements of third party content appear on page within text, which constitute an extension of this copyright page.

Unless otherwise indicated herein, any third-party trademarks that may appear in this work are the property of their respective owners and any references to third-party trademarks, logos or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc. or its affiliates, authors, licensees or distributors.

Library of Congress Cataloging-in-Publication Data

Popham, W. James.

Classroom assessment : what teachers need to know / W. James Popham Professor Emeritus, University of California, Los Angeles.—Eighth edition.

pages cm

ISBN 978-0-13-405386-8

1. Educational tests and measurements. 2. Examinations—Design and construction.

3. Examinations—Validity. I. Title.

LB3051.P6142 2017

371.260973—dc23

2015025822

10 9 8 7 6 5 4 3 2 1

PEARSON

ISBN 10: 0-13-405386-9

ISBN 13: 978-0-13-405386-8

*For Sarah:
Former Teacher, Former Principal,
Current Colleague, Current and Future Spouse*

about the author

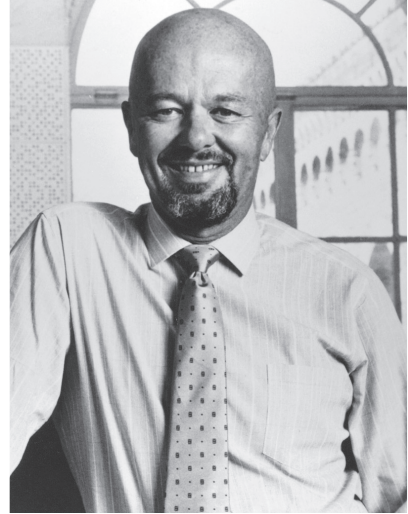
W. James Popham has spent the bulk of his educational career as a teacher. His first teaching assignment, for example, was in a small eastern Oregon high school where he taught English and social studies while serving as yearbook advisor, class sponsor, and unpaid tennis coach. The recompense meshed well with the quality of his coaching.

Most of Dr. Popham's teaching career took place at UCLA where, for nearly 30 years, he taught courses in instructional methods for prospective teachers as well as courses in evaluation and measurement for graduate students. At UCLA he won several distinguished teaching awards. In January 2000, he was recognized by *UCLA Today* as one of the university's top 20 professors of the twentieth century. (He notes that the twentieth century was a full-length century, unlike the current abbreviated one.) In 1992, he took early retirement from UCLA upon learning that emeritus professors received free parking.

Because at UCLA he was acutely aware of the perishability of professors who failed to publish, he spent his nonteaching hours affixing words to paper. The result: over 30 books, 200 journal articles, 50 research reports, and 175 papers presented before research societies. Although not noted in his official vita, while at UCLA he also authored 1,426 grocery lists.

His most recent books are *Transformative Assessment* (2008); *Instruction That Measures Up* (2009); *Transformative Assessment in Action* (2011, ASCD); *Mastering Assessment* (2011, Pearson); *Unlearned Lessons* (2009, Harvard Education Press); *Everything School Leaders Need to Know About Assessment* (2010); and *Evaluating America's Teachers: Mission Possible?* (2013, Corwin). He encourages purchase of these books because he regards their semi-annual royalties as psychologically reassuring.

In 1968, Dr. Popham established IOX Assessment Associates, an R&D group that formerly created statewide student achievement tests for a dozen states. He has personally passed all of those tests, largely because of his unlimited access to the tests' answer keys.



preface

Educational Assessment in Flux

Perhaps you have heard the ancient Chinese curse, *May you live in an interesting time!* Perhaps you haven't.

Well, I can definitely tell you where and when I first heard this curse—and how puzzled I was by its meaning. The year was 1961, and I was a rookie assistant professor at San Francisco State College. A campuswide speech was to be delivered by Robert Maynard Hutchins, an educational celebrity of that era. Hutchins was the founder of the Great Books Movement and had been the youngest-ever chancellor of the University of Chicago.

It was a simply marvelous speech—so fine, in fact, that I subsequently obtained an audiotape of the address and played it often in my classes. Hutchins opened his address with the following sentence: “Perhaps you have heard the ancient Chinese curse, ‘*May you live in an interesting time!*’”

As I indicated, upon hearing Hutchins's first sentence, I was immediately perplexed by the meaning of this “curse” that I'd never heard before. After all, if the time in which one lives is “interesting,” this would seem to be a positive—not a negative. What's interesting is typically better than what's dull. But then, as Hutchins continued, he pointed out that an “interesting time” invariably involves *changes*. Indeed, the more profound the changes, the more “interesting” the time. And changes, at least for most of us, cause *discomfort*. We must accommodate to what's new. Routine, experience-honed approaches no longer work. New, “interesting” times simply bristle with uncertainty. Hutchins was warning his audience that education in the United States was entering an era of unprecedented change and, as a consequence, U.S. educators should clearly regard themselves as consummately cursed.

Well, if you look at what's taking place these days regarding this nation's *educational assessment*, you'll quickly conclude that we are smack in the middle of what is, most certainly, an especially “interesting time.” To illustrate, as the revision of the classroom-assessment textbook you're currently reading was nearing the finish line, U.S. educators were still perplexed about how to deal with the Common Core State Standards, a set of curricular aims that, chiefly because of federal financial incentives, had been adopted by many, but not all, states. Beyond that, two different consortia of states that had set out to create “next generation” assessments to measure students' mastery of those Common Core State Standards were beginning to assess students in many parts of the nation. Because those new assessments measured students' mastery of what were widely regarded as more challenging curricular goals, students' performances on the new consortium-built

tests were predictably lower than students' scores on other tests in previous years. Finally, Congressional reauthorization of the Elementary and Secondary Education Act of 1965 (ESEA) was long overdue and, although the current version of that federal statute—the No Child Left Behind Act—was clearly in need of a serious re-do, realistic estimates of when Congress might finally get around to reauthorizing this law ranged widely.

Yet, because American educators are now coping with assessment events that, in concert, constitute a truly “interesting time,” I hope you will be patient with a textbook author who must come up with a sensible eighth edition of a classroom assessment textbook in such a setting. Unlike any of its previous versions, this rendition of *Classroom Assessment* was written in the midst of kaleidoscopic confusion about what sorts of assessments are around the corner and what roles those assessments are going to play in our schools. This eighth edition of *Classroom Assessment*, I fear, is apt to have been afflicted by a curse—probably of Chinese origin.

Formative Assessment in Action

This most recent revision of *Classroom Assessment*, however, is fundamentally different than its seven predecessors, and that difference stems directly from a phone call I received a couple of years ago from my editor, Kevin Davis. It was a phone call that seriously mucked up the next two years of my life.

My editor opened with an upbeat inquiry: “We’d like the next edition of *Classroom Assessment* to be digital. Do you have any problem with that?” Having always believed in electricity, I had no problem with it at all—especially after Kevin described what the possibilities were for a digitally presented textbook. The more I considered this new publication strategy, the more enthused I became—for I realized that a digitized eighth edition would make it possible for the book to be revised so it could become *an actual incarnation of the formative-assessment process*. And, even though it took much, much more work to revise the book, that’s precisely what this eighth edition of *Classroom Assessment* is—an exemplification of how the formative-assessment process can support learning.

When formative assessment is present, students periodically provide assessment-elicited evidence that teachers can use to tell if adjustments are needed in their instruction or that students can use to tell if adjustments are needed how they’re trying to learn. Happily, we now have an impressive collection of empirical evidence indicating the formative-assessment process is a research-ratified way of helping both teachers and students become more effective. When you read this eighth edition as a student, you should be on the receiving end of formative assessment’s payoffs.

Thus, at the outset of each of this new edition’s 16 chapters, a *chief chapter outcome* has been identified. This outcome is the most significant understanding

you should gain from reading the chapter. Then, at the chapter's conclusion, two different self-tests are provided for you to determine how well you have mastered that chapter's chief outcome. (One of these mastery checks employs selected-response items, and one calls on you to generate a short constructed-response essay.) Having completed a chapter's two mastery checks, and decided how well you achieved the chapter's chief outcome, you then have three options available to you. If you judged that you demonstrated satisfactory mastery of the chapter's chief outcome, you could simply move ahead to the book's next chapter. If, however, you performed satisfactorily on the chapter's two mastery checks—and also found the chapter's content to be particularly fascinating—you could consult the array of digital options representing a *deeper dig* into one or more chapter-related issues. Finally, if your performance on one or both of the end-of-chapter mastery checks suggests that you need to spend more time in pursuit of the chapter's chief outcome, then you can refer to the digitally presented segments constituting *another take* regarding the chapter's content. In this book's application of the formative-assessment process, as in similar successful applications of formative assessment, it is the learner (you) who often uses assessment results to determine what should be done next.

Hopefully, reconfiguring the eighth edition of *Classroom Assessment* into a model of the formative-assessment process will have two clear-cut benefits. First, if you use each chapter's mastery checks to decide how well you have attained the chapter's chief intended outcome, you can then benefit by making your next activity an evidence-informed choice rather than a data-free conjecture. Second, because this new edition represents an attempt to adopt, insofar as a textbook-only approach to formative assessment permits—it is anticipated that this modeling may incline at least a portion of its teacher-readers to employ some variation of formative assessment in their own classes.

Classroom Assessment's embrace of digitization, as indicated earlier, required a far more laborious revision of this book than I had ever undertaken with previous editions. Yet, I am more optimistic about this new, formatively oriented revision than I have ever been. I truly hope that you not only like the eighth edition's approach, but that its digitized dazzle works as advertised.

New to This Edition

Given the uncertain nature of tomorrow's high-stakes assessments and the legal requirements regarding how they must be used, I've tried to comb through the book to make sure that heretofore unequivocal statements about state-level assessments and state-level accountability laws have been suitably softened with dollops of uncertainty. The varnish-free truth is that, as I was readying this revision, education leaders in a good many states had told me, off the record, they simply had no idea about what sorts of state-level assessment programs would be present during

the next few years. What, a few years earlier, had been uniform adulation of the *Common Core State Standards* is currently a much more mixed bag of support or disdain. Similarly, early optimism had been widely registered regarding the two emerging sets of consortium-built tests that would allow American educators to gauge students' mastery of challenging "college and career-ready" content. More recently, however, doubts have been voiced by some educators regarding these "next generation" assessments.

In mid-2014, three of our nation's organizations most concerned with educational measurement released a genuinely significant document—namely, the *Standards for Educational and Psychological Testing*. The standards (guidelines) included therein describe how education tests should be constructed, evaluated, and used. Because the standards in this publication are often relied on to resolve assessment-related courtroom litigation, the tenets set forth in this volume will have a profound impact on the way U.S. educational testing takes place in the coming years. The most important of these new standards have been incorporated into all relevant sections of *Classroom Assessment*. Given today's pervasive pressure to boost students' test scores, a number of test publishers are hawking what they characterize as "diagnostic tests." To help teachers understand what's necessary for a diagnostic test to make a meaningful contribution to instruction, a new section on *instructionally diagnostic testing* has been included in Chapter 13. Readers who understand what's needed in a test that is truly diagnostic will be better able to choose among an ever-expanding array of supposedly diagnostic assessments. Many of today's allegedly useful diagnostic tests should, regrettably, be sent to a shredder. The new section on instructionally diagnostic tests will help readers evaluate the merits of commercial or locally developed "diagnostic" assessments.

The "Legacy" Elements of the Book

A number of features in this current edition, as might be expected, are inherited from previous renditions of the book. As in past editions, this new version is attentive to the *instructional* payoffs of well-designed classroom tests. Whenever I can, I've attempted to highlight the implications that testing has on teaching. In each chapter, there's even a section titled *But What Does This Have to Do with Teaching?* In those instances, I spell out the instructional implications of a chapter's content.

All chapters contain a description of fictitious classroom teachers who must make decisions related to one or more of the topics treated in the chapter. These are called *Decision Time* features, and each one concludes with a decision that needs to be made by the teacher being described. Readers are then asked to put themselves in the place of the teacher and decide how they would proceed. These vignettes are intended to set out a series of practical problems squarely on a reader's plate, and then see how readers might munch on them.

Considerable attention is also given in this eighth edition to an important audience for teachers—namely, *parents*. Teachers often need to explain to parents such assessment-related topics as why a child scored a certain way on a standardized test or how a teacher’s classroom exams are related to the grades a child receives. So, in every chapter, you’ll find a feature titled *Parent Talk*. In these Parent Talk segments, I’ve described a situation in which a teacher needs to explain something about assessment to parents. I then indicate what I would have said to the parents if I had been the teacher in that situation. What I hope readers will do, then, is decide how *they* would have responded had they been placed in this same situation. In fact, readers might even say *aloud* (in private, if they have any sense) what they’d tell parents. Teachers who can talk sensibly to parents about assessment-related concerns will find they’re able to establish more effective rapport with parents. Such teachers will get along far better with parents than will teachers who convey to parents the idea that assessment is an exotic measurement mystery, well beyond the perceptive powers of mere parents.

I realize all too well that most readers are not likely to regard the content of a book on assessment as enthralling. For myself, if I had the choice between whether I would read an assessment book or a cracking good espionage novel, I’d shove the assessment book aside in a millisecond. In recognition of a reader’s likely response to the book’s content, I’ve tried to lighten the load with cartoons and an occasional dash of levity. If readers don’t find the light style to be acceptable, they are encouraged to frown during the funny parts.

MyEducationLab[®]: Digitalization Does Its Dance

Now, to describe how the digital features of the eighth edition work, here are the basics of that operation. In the MyEdLab for this book, at the end of each chapter you will find three links:

- MyEdLab: Selected-Response Check of Outcome Mastery
- MyEdLab: Constructed-Response Check of Outcome Mastery
- MyEdLab: Learning Outcome Mastery Determination

The first two links take you to interactive mastery checks (one selected-response and the other constructed-response). Both these mastery checks contain feedback that enable you to reach a judgment regarding whether you have mastered the chapter’s chief intended outcome. The following provide examples of the two types of Outcome Mastery checks.

Determining Your Outcome Mastery

The importance of validity to educational assessment really cannot be over-emphasized. It is, unarguably, the most influential concept in all of educational testing. Teachers need to know what valid sort of evidence is needed to support test-based purposes. Let's look, then, at the major outcome after you completed this chapter:

A sufficiently deep understanding of an essential nature can be explained, its essential nature, and the most appropriate kinds of validity evidence can be selected for specific uses of educational tests

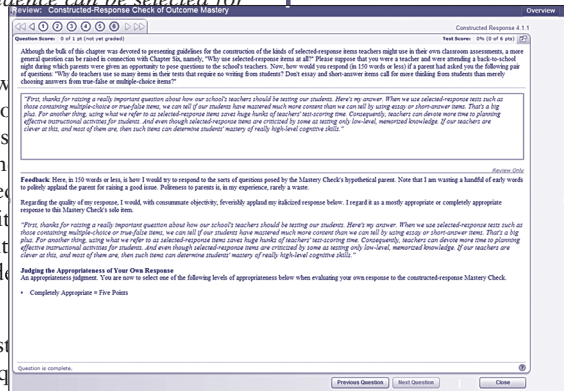
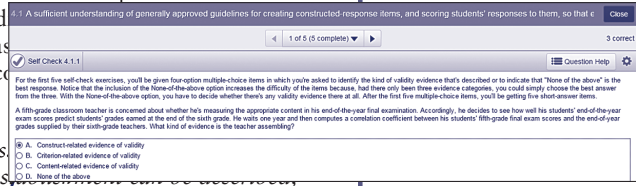
With the completion of the chapter, you should now describe what validity is, (2) explain how validity can be used, and (3) choose suitable validity evidence for diverse uses. In essence, you should have the ability to display a fundamental validity by being able to carry out three related tasks regarding validity. Using nonfancy language, it is to understand what's going on when people talk about validity, to be well that you can comfortably do all three things in order to achieve the chief intended learning outcome.

Complete both the Selected-Response and the Constructed-Response checks and think about the feedback you receive for each quiz.

MyEdLab [Selected-Response Check of Outcome Mastery](#)
MyEdLab [Constructed-Response Check of Outcome Mastery](#)

After completing both quizzes, go to the Learning Outcome Mastery Determination, where you will decide whether you've mastered the chapter's learning outcome or whether you need further study.

MyEdLab [Learning Outcome Mastery Determination](#)



After completing both mastery checks, you will be instructed to go to the third link—Learning Outcome Mastery Determination, where you will decide whether you have mastered the chapter's intended outcome or whether you need further study. If you believe that you have not mastered the chapter's chief outcomes, you are not directed to re-read the chapter. Instead, new explanations are provided in *Another Take* segments. If, based on how well you did on the mastery checks, you believe you have mastered the chapter's chief intended outcome, you will be given an opportunity to extend your knowledge of the chapter's contents by exploring material that probes further into the assessment concepts presented in the chapter. These materials are called *Deeper Digs* segments.

Here, of course, is where the heart of the formative-assessment process beats. You will be collecting your own evidence of content mastery, then deciding what the next step in your learning efforts should be. If you think you performed well on a chapter's two mastery checks, are you obliged only to consider the Deeper Dig segments—and never dip into the *Another Take* segment? Of course not. Similarly, you can surely probe any *Deeper Dig* segments even if you didn't do all that well on a given chapter's mastery checks. In general, the decisions you make about your own learning activities at the close of a chapter are apt to be more defensible if you base those decisions on your assessed mastery of the chapter's most significant learning outcome. Nonetheless, in this digital age—and with this digitized new edition—the choices are clearly yours.

Why This Book Will Help Today's Teachers

Teachers these days who don't recognize that educational assessment impinges on their work are teachers in serious need of impingement therapy. Rarely, indeed, does a day go by in the Monday-through-Friday life of today's teachers when testing does not have an impact on one or more of their classroom decisions. It was not always this way.

Eons ago, I was a high school teacher in eastern Oregon. (It was so long ago that my friends contend Oregon must have been a territory rather than a state.) Way back then, we administered standardized achievement tests in our classes. However, students' scores on those tests made no difference in how we taught. Pressure to raise our students' scores on those achievement exams was nonexistent. We taught pretty much as we saw fit. But, of course, the world of education is different today—much different.

And even before those teaching days, when I was preparing to be a teacher, little attention was given to testing. In truth, the only time my professors actually taught us about educational tests was when, during an educational psychology class, we spent an entire week on the making and massaging of multiple-choice items. My fellow prospective teachers and I were not being prepared for educational assessment because, back then, educational assessment truly did not have an impact on teachers' decision-making.

But today, educational tests certainly make a difference regarding what currently takes place in our classrooms. For openers, today's teachers find themselves directly in the cross-hairs of some heavy-duty accountability artillery aimed at evaluating schools *and teachers* according to students' scores on accountability tests. A school's staff can be "restructured" or a school can be completely shut down if its students don't perform well enough on externally administered accountability exams. Teachers can be tossed. It is a scary time.

Second, during the last two decades, growing numbers of educators have learned that the skillful use of classroom testing can make huge differences in how

well students learn. Classroom assessment, if employed *formatively*, can dramatically increase the effectiveness of a teacher's *teaching*. And yet, sadly, we rarely see more than token use of classroom assessment in the way that research clearly tells us will benefit students.

For both of those reasons, then, every experienced teacher and every teacher-in-training need to master the essentials of educational assessment. And that's why this book was first written and then revised so frequently. Its title, *Classroom Assessment: What Teachers Need to Know*, captures the book's intent. Readers won't be asked to learn any nice-to-know exotics about educational measurement. No, what's contained in this book is the stuff today's teachers *need to know* if they are going to be first-rate professionals.

Acknowledgments

To my colleagues who offered recommendations about how to improve this latest edition, I offer my sincere appreciation: Deborah Bennett, Purdue University; Jean Ragin, Coppin State University; Marilyn Roseman, Mount Aloysius College; and Christy Tirrell-Corbin, University of Maryland.

I am especially grateful to Dolly Bulquerin, great friend and word-processor nonpareil, who prepared the very first edition of *Classroom Assessment*. She then helped transform that first edition into second, third, fourth, fifth, sixth, and seventh versions. Now, once more, she has helped me birth this exciting new eighth edition. I suspect those efforts clearly have established Dolly as a multi-edition midwife.

W. J. P.

brief contents

1	Why Do Teachers Need to Know about Assessment?	1
2	Deciding What to Assess	32
3	Reliability of Assessment	72
4	Validity	94
5	Fairness	124
6	Selected-Response Tests	150
7	Constructed-Response Tests	174
8	Performance Assessment	195
9	Portfolio Assessment	220
10	Affective Assessment	236
11	Improving Teacher-Developed Assessments	257
12	Formative Assessment	274
13	Making Sense Out of Standardized Test Scores	307
14	Appropriate and Inappropriate Test-Preparation Practices	339
15	The Evaluation of Instruction	353
16	Assessment-Based Grading	385
	Glossary	406
	Index	412

1

Why Do Teachers Need to Know about Assessment? 1

Chief Chapter Outcome 1

Federal Laws Rule 2

Assessment versus Testing 9

Why Should Teachers Know about Assessment? Yesteryear's Answers 12

Why Should Teachers Know about Assessment? Today's Answers 17

What Do Classroom Teachers Really Need to Know about Assessment? 23

Chapter Summary 29

Determining Your Outcome Mastery 30

References 31

2

Deciding What to Assess 32

Chief Chapter Outcome 32

What to Assess 32

A Collection of What-to-Assess Considerations 59

What Do Classroom Teachers Really Need to Know about What to Assess? 66

A Three-Chapter Preview 67

Chapter Summary 69

Determining Your Outcome Mastery 69

References 70

3

Reliability of Assessment 72

Chief Chapter Outcome 72

Test-Retest Reliability Evidence 76

Alternate-Form Reliability Evidence 79

Internal Consistency Reliability Evidence 81

Three Coins in the Reliability/Precision Fountain	84
The Standard Error of Measurement	85
What Do Classroom Teachers Really Need to Know about Reliability/Precision?	88
Chapter Summary	91
Determining Your Outcome Mastery	92
References	92
Endnotes	93

4 **Validity** 94

Chief Chapter Outcome	94
A Quest for Defensible Interpretations	94
Validity Evidence	99
Validity Evidence Based on Test Content	102
Validity Evidence Based on Response Processes	109
Validity Evidence Based on a Test's Internal Structure	110
Validity Evidence Based on Relations to Other Variables	112
Sanctioned and Unsanctioned Labels for Validity Evidence	117
The Relationship between Reliability and Validity	118
What Do Classroom Teachers Really Need to Know about Validity?	119
Chapter Summary	121
Determining Your Outcome Mastery	122
References	123

5 **Fairness** 124

Chief Chapter Outcome	124
The Nature of Assessment Bias	125
Disparate Impact and Assessment Bias	128
Judgmental Approaches	130
Empirical Approaches	133
Bias Detection in the Classroom	133
Assessing Students with Disabilities and English Language Learners	136
What Do Classroom Teachers Really Need to Know about Fairness?	145

Chapter Summary 147
Determining Your Outcome Mastery 148
References 149
Endnote 149

6 Selected-Response Tests 150

Chief Chapter Outcome 150
Expanding Electronic Options 151
Ten (Divided by Two) Item-Writing Commandments 151
Binary-Choice Items 156
Multiple Binary-Choice Items 159
Multiple-Choice Items 161
Matching Items 166
What Do Classroom Teachers Really Need to Know about Selected-Response Tests? 170
Chapter Summary 172
Determining Your Outcome Mastery 172
References 173

7 Constructed-Response Tests 174

Chief Chapter Outcome 174
Short-Answer Items 175
Essay Items: Development 178
Essay Items: Scoring Students' Responses 183
What Do Classroom Teachers Really Need to Know about Constructed-Response Tests? 191
Chapter Summary 193
Determining Your Outcome Mastery 193
References 194

8**Performance Assessment 195**

- Chief Chapter Outcome 195
- What Is a Performance Test? 195
- Identifying Suitable Tasks for Performance Assessment 198
- The Role of Rubrics 203
- An Illustrative Performance-Test Task 205
- Wretched and the Rapturous Rubrics 205
- Ratings and Observations 213
- Sources of Error in Scoring Student Performances 214
- What Do Classroom Teachers Really Need to Know about Performance Assessment? 216
 - Chapter Summary 217
 - Determining Your Outcome Mastery 218
 - References 219

9**Portfolio Assessment 220**

- Chief Chapter Outcome 220
- Classroom Portfolio Assessment versus Large-Scale Portfolio Assessment 221
- Key Ingredients in Classroom Portfolio Assessment 224
- The Pros and Cons of Portfolio Assessment 229
- What Do Classroom Teachers Really Need to Know about Portfolio Assessment? 232
 - Chapter Summary 234
 - Determining Your Outcome Mastery 234
 - References 235

10**Affective Assessment 236**

- Chief Chapter Outcome 236
- Why Assess Affect? 236
- The Other Side of the Argument 239
- Which Affective Variables Should Be Assessed? 240
- How Should Affect Be Assessed in Classrooms? 243

Self-Report Assessment 244
What Kinds of Inferences Are at Stake in Affective Assessment? 252
What Do Classroom Teachers Really Need to Know about Affective Assessment? 254
 Chapter Summary 255
 Determining Your Outcome Mastery 255
 References 256

11 Improving Teacher-Developed Assessments 257

Chief Chapter Outcome 257
Judgmentally Based Improvement Procedures 258
Empirically Based Improvement Procedures 262
Item Analysis for Criterion-Referenced Measurement 268
What Do Classroom Teachers Really Need to Know about
 Improving Their Assessments? 271
 Chapter Summary 272
 Determining Your Outcome Mastery 272
 References 273

12 Formative Assessment 274

Chief Chapter Outcome 274
Assessment That Transforms Teaching 274
Formative Assessment Research Support 281
Learning Progressions as Frameworks 284
Alternative Implementation Approaches 290
Why the Delay? 299
What Do Classroom Teachers Really Need to Know about
 Formative Assessment? 302
 Chapter Summary 303
 Determining Your Outcome Mastery 304
 References 305

13**Making Sense Out of Standardized Test Scores 307**

- Chief Chapter Outcome 307
- Standardized Tests 308
- Group-Focused Test Interpretation 310
- Individual Student Test Interpretation 312
- Instructionally Diagnostic Tests 324
- The SAT and the ACT: Three-Letter, High-Import Exams 330
- What Do Classroom Teachers Really Need to Know about Interpreting Standardized Test Scores? 335
 - Chapter Summary 336
 - Determining Your Outcome Mastery 337
 - References 338

14**Appropriate and Inappropriate Test-Preparation Practices 339**

- Chief Chapter Outcome 339
- High-Stakes Assessment Arrives 339
- Assessment Results as Inference Illuminators 340
- Two Evaluative Guidelines 342
- Five Test-Preparation Practices 344
- Applying the Two Guidelines 345
- What about “Teaching to the Test”? 348
- What Do Classroom Teachers Really Need to Know about Test-Preparation Practices? 350
 - Chapter Summary 351
 - Determining Your Outcome Mastery 351
 - References 352

15**The Evaluation of Instruction 353**

- Chief Chapter Outcome 353
- A Focus on Consequences 355

Classroom Assessment Evidence 358
Evidence from External Accountability Tests 363
What's a Teacher to Do? 378
What Do Classroom Teachers Really Need to Know about
Assessment-Based Evaluation of Teaching? 380
Chapter Summary 382
Determining Your Outcome Mastery 382
References 383

16 **Assessment-Based Grading** 385

Chief Chapter Outcome 385
The Purpose of Grading 385
Grade-Giving Specifics 392
What Do Classroom Teachers Really Need to Know about
Assessment-Based Grading of Students? 402
Chapter Summary 404
Determining Your Outcome Mastery 404
References 405

Glossary 406

Index 412

Why Do Teachers Need to Know about Assessment?

CHIEF CHAPTER OUTCOME

An understanding of why it is that four traditional and three recent reasons for educators to assess students should dispose teachers to learn more about the fundamentals of educational assessment

Teachers teach students. That hardly constitutes a breakthrough insight. Just as preachers preach and flyers fly—teachers teach. That’s why they’re called teachers.

But what is a bit less obvious is that most teachers teach because they *like* to teach. Primary teachers like to teach little people. High school teachers like to teach bigger people. Most high school teachers also like to teach about a particular subject matter. (Have you ever seen how mathematics teachers’ eyes get misty when they introduce their students to the raptures of the Pythagorean theorem?) Yes, most teachers love to teach. It is because they enjoy what

they do that they waded through a medley of preservice teacher education courses, conquered the challenges of student teaching, and hopped the myriad hurdles of the certification process. Teachers overcame these obstacles in order to earn annual salaries that, particularly during the first few years, are laughably low. Yes, there’s little doubt that teachers enjoy teaching.

Although teachers like to *teach*, they rarely like to *test*. Yet, here you are—beginning a book about testing. How can I, the author, ever entice you, the reader, to become interested in testing when your heart has already been given to teaching? The answer is really quite straightforward. Teachers who can test well will be better teachers. Effective testing will enhance a teacher’s instructional effectiveness. Really!

If you’re willing to suspend any preconceptions about testing while you’re reading this book, particularly any negative ones, I’ll make a pledge to you. If you tackle this text with even half the enthusiasm you might bring to a teaching assignment, I promise you’ll discover how *testing will make you a much better teacher*. And, because I’ve been a teacher for over 50 years, it’s a promise I’ll keep. Teachers definitely should not break promises to teachers. Teachers’ promises to administrators, on the other hand, should be regarded as eminently renegotiable.

But before I attempt to convince you, ever so subtly, that testing can be a boon to teaching, I want you to get a fix on your own *current* views about

educational testing. And, because this is a book about testing, what better way to have you learn about those attitudes than to have you take a self-test I devised just for readers of this book?

So, on the adjacent page (it's on the *right* from where you're currently reading!) you'll find a brief self-test similar to the ones you've surely encountered in many widely read magazines. I saw one such self-test in a health magazine recently. It was entitled "How Long Will You Live? A Self-Test." Frankly, I was afraid to try it. As one gets older, one becomes more cautious.

But you have nothing to fear by taking the self-test I've whipped up for you. To emphasize its brevity, I have entitled it "A Terse Self-Test about Testing." It is an example of an attitudinal inventory. Later, in Chapter 10, you'll learn more about attitudinal inventories. But for now, please take a crack at page 3's teensy self-test. The way to interpret your responses is given as a footnote at the bottom of page 4.

FEDERAL LAWS RULE

Anyone who has completed even an introductory course in U.S. Government knows that while state laws can overturn the laws enacted by local communities, federal laws can overturn state laws. When it comes to the art of overturning, federal folks hold all the trump cards.

Any consideration of educational testing these days cannot be sensibly undertaken without understanding the nature of whatever assessment-related federal laws are on the books. When I began working on this eighth edition of *Classroom Assessment*, the most significant education-related federal law then in place was the No Child Left Behind Act (NCLB). Although NCLB has exercised considerable influence on the way U.S. teachers tested and taught in their classrooms, the law elicited intense criticism from many quarters. Moreover, NCLB was supposed to be revised sometime during 2009 or, at the latest, 2010. Yet, by early-to-mid-2015, genuinely serious movement to revise NCLB had not yet surfaced in the U.S. Congress. Accordingly, because this eighth edition of the book would most likely be completed before a successor-law to NCLB had been enacted, it seemed silly to speculate about what the key assessment-related features of such a yet-unwritten law might be.

Instead, very briefly, I want to describe the background of the most pivotal federal legislation that, in one form or another, will surely have an impact on the way teachers are obliged to think about educational testing. Hopefully, based on that familiarity, you will then be more easily able to learn about the particulars of any federal law bearing directly on how students' achievements are supposed to be assessed. All educators will definitely need to attend to those assessment-related particulars.

By all odds, the most significant federal statute influencing U.S. educational testing was the Elementary and Secondary Education Act (ESEA) of 1965.



A Terse Self-Test about Testing

Directions: For each of the statements below, use the following answer key to indicate how you react to the statement:

SA = *Strongly Agree*

A = *Agree*

U = *Uncertain*

D = *Disagree*

SD = *Strongly Disagree*

There are no right or wrong answers, so please answer frankly by circling the appropriate response for each statement.

- | | | | | | |
|---|----|---|---|---|----|
| 1. The chief reason that teachers should give classroom tests is to determine students' grades. | SA | A | U | D | SD |
| 2. Teachers should typically plan instruction that focuses on the skills or knowledge represented by a test. | SA | A | U | D | SD |
| 3. In their classroom tests, teachers should only use items that can be scored objectively. | SA | A | U | D | SD |
| 4. There are other legitimate indicators of a teacher's instructional effectiveness besides students' test scores. | SA | A | U | D | SD |
| 5. A teacher has no business measuring students' confidence in their ability to do schoolwork. | SA | A | U | D | SD |
| 6. Today's nationally standardized achievement tests should never be used to supply evidence about how well teachers are instructing children. | SA | A | U | D | SD |
| 7. Teachers really don't need to determine the reliability of their own classroom tests. | SA | A | U | D | SD |
| 8. It is impossible to judge the quality of students' written compositions with any meaningful accuracy. | SA | A | U | D | SD |
| 9. The enormous pressure to boost students' scores on important tests permits teachers to employ almost any sort of score-improvement preparation activities. | SA | A | U | D | SD |
| 10. Significant classroom tests should typically be built before a teacher plans instruction. | SA | A | U | D | SD |

Enacted as a key component of President Lyndon B. Johnson’s “great society,” ESEA set out to provide a more appropriate education for historically underserved student groups such as students who were economically disadvantaged. Over the years (actually, every two to eight years), ESEA was periodically reauthorized with, sometimes, serious shifts in its assessment provisions. No Child Left Behind, for instance, the eighth reauthorization of 1965’s ESEA, contained some significant alterations to that law’s testing requirements. In the earliest incarnations of ESEA, educational assessments were focused on evaluating the progress made by those statutorily designated underserved groups—for example, minority students. However, in the reauthorization immediately preceding NCLB, the reauthorization enacted in 1994, the assessment of *all* students rather than statute-designated underserved groups was required. Clearly, this was a change of considerable importance.

Because the responsibility for education is not identified as a federal responsibility in the United States Constitution, U.S. education has historically been seen as a state rather than federal responsibility. Thus, prior to the 1994 incarnation of ESEA, known as the Improving America’s Schools Act (IASA), states were relatively free to carry out whatever sorts of educational assessments they thought appropriate, with the chief exception being the assessment of those students being educated, at least in part, via federal dollars dispensed by the then-operative version of ESEA. But in 1994’s IASA, that game changed. When a state took IASA dollars, this state agreed to assess the achievement of all its students in several IASA-designated grade ranges. And when NCLB was signed into law by President George W. Bush on January 8, 2002, the assessment of *all* students became more emphatic by far. Students in twice as many grade levels (grade *levels*, not grade ranges) were to be assessed—even though many of those students were not on the receiving end of federal dollars.

Moreover, whereas in the IASA statute, federal oversight of state-level testing of students in certain grade ranges was fairly light, NCLB’s controls over the testing of more than twice as many students assessed under that law was not light but, instead, quite tight indeed. In short, the most recent two versions of ESEA (IASA and NCLB) embodied increasingly stringent requirements regarding which students were to be tested and how this testing was to be done. While the dominant function of IASA and NCLB was to be accountability—that is, the identification of which schools and districts were doing a satisfactory instructional job—certain specifics of those laws make a real difference in how teachers need to think about educational assessment.

Self-Test Interpretation Guide: For statements 2, 4, 6, 7, and 10, use the following scoring key: SA = 5, A = 4, U = 3, D = 2, and SD = 1. For statements 1, 3, 5, 8, and 9, use the following scoring key: SA = 1, A = 2, U = 3, D = 4, and SD = 5. The highest possible total score is 50; the lowest possible total score is 10. The higher your total score, the more sensible is your view of educational testing. After finishing this book, you might wish to retake this terse self-test (without looking at your earlier answers, of course). If you come up with a postbook score that’s substantially *lower* than your prebook score, you and I should *both* be worried.

Assuming that the next reauthorization of ESEA will not be in place at the point when this book must go into actual production, I entreat you to become knowledgeable about the assessment-related aspects of a reauthorized ESEA. Although there may be numerous features of such a law that can have an impact on the way teachers teach, it is almost certain that the assessment-related provisions of such a law will have great impact, if not the greatest impact, on how a teacher needs to think about *instruction*. In the realm of educational assessment, federal laws tend to rule. That's because federal legislators craft their statutes so that unless a state's officials comply with a federal statute's ground rules, that state must forego receipt of substantial federal dollars. The history of American public education is, as you might guess, not replete with instances wherein state authorities turned down federal dollars.

Interestingly, when the NCLB statute experienced its decade-old anniversary, the federal government's stance regarding how best to foster state and local accountability initiatives had shifted considerably. The early years of NCLB's existence had been marked by the threat of penalties for low-performing schools. However, after President Barack Obama's administration had taken office, federal officials soon set out meaningful financial incentives for states who subscribed to the U.S. Department of Education's accountability preferences. As usual, state education officials responded predictably to the lure of federal largesse. In essence, then, federal implementation of ESEA had shifted—in just a few years—from the stick to the carrot.

Arrival of the Common Core State Standards

One of the most salient of these carrot-induced shifts in state education policies was associated with the adoption, by all but a few states, of a set of identical curricular aims: the *Common Core State Standards (CCSS)* in English language arts (ELA) and mathematics. Because states that adopted these identical curricular aims became eligible for receipt of substantial federal subsidies, and because a markedly slowing national economy found most states facing serious fiscal shortfalls, educators soon saw almost all states accepting—as their state's official curricular aims—the CCSS in ELA and math. This event, almost unthinkable just a few years earlier, was near certain to have a substantial impact on the instructional and assessment practices of the nation's public school teachers in the coming years. Although educational authorities in the vast majority of U.S. states have adopted the CCSS as the official curricular aims in their state, since the early months of what seemed, in retrospect, to be an “adoption orgy” with almost all states hopping aboard the CCSS bandwagon, educational leaders in some states have now hopped off. The reasons for this turnaround in educational policy during a relatively brief span of years are several. Most often, we have seen officials in some states arguing that the CCSS represents a federal intrusion into the education of our young—historically an enterprise undertaken by states, not the federal government. Thus, backpedaling by CCSS states regarding adoption of the CCSS appears to be based more on

political than educational rationales. In at least a few states, however, the educational leaders of those states (or, sometimes, the members of a state's legislature) found themselves in disagreement with certain of the curricular emphases of the CCSS. Although, as this edition of *Classroom Assessment* headed off happily to the publishers, the final number of U.S. states adopting the CCSS was uncertain, a good many states have either adopted the aims embodied in the CCSS or have made only slight modifications in the CCSS curricular goals, then adopted those substantively similar curricular aims. To be sure, in certain states we have seen truly acrimonious disputes among educational policymakers regarding their state's acceptance of the curricular aspirations embodied in the CCSS.

Let's look, ever so briefly, at what these curricular aims are—with a definite commitment to return in the next chapter for a deeper dip into the viscera of the CCSS. In Chapter 2, you will see how the two sets of curricular aims identified in the CCSS are organized, as well as hear what some of the developers of those state standards were hoping to accomplish.

Let's be clear about what the *Common Core State Standards* are. They represent the curricular outcomes sought for the nation's students—that is, the knowledge and cognitive skills students are supposed to acquire in school. Because NCLB had allowed each state to select its own curricular aims (that is, content standards), its own tests to assess students' mastery of those aims, and its own cut-scores (that is, achievement standards) to signify students' mastery of those curricular aims, making sense out of the NCLB-spawned accountability picture in U.S. public schools was almost impossible. In an effort to rectify this chaotic situation, the Council of Chief State School Officers (CCSSO) and the National Governors Association (NGA) Center for Best Practices set out in late 2009 to provide a more suitable set of curricular targets for the nation's schools. The CCSSO is the organization of the state officials, elected or appointed, who head each state's public schools. The NGA performs a comparable function for the nation's governors.

On June 2, 2010, the CCSSO and the NGA released the *Common Core State Standards for English Language Arts and Mathematics* (National Governors Association, 2010). As noted earlier, many states have accepted these standards—these “expectations for student knowledge and skills that high school graduates need to master to succeed in college and careers.” Given the long-standing reluctance of state education officials to abandon “local control” over important educational decisions such as curricular outcomes for students, the widespread adoption of the CCSS was genuinely astonishing. In essentially a single year, the CCSSO and the NGA crafted sets of national mathematics and ELA curricular aims that seem sufficiently defensible so that all but a few states soon hopped aboard the CCSS Express.

The widespread and remarkably rapid adoption of the CCSS by so many states, however, did not take place merely because of the merits of a more uniform set of curricular targets for America. The recent role of philanthropic organizations in nurturing such significant changes in U.S. education is now being better understood.

In the June 7, 2014, issue of *The Washington Post*, Lyndsey Layton reports that a major player in the adoption of the CCSS was the Bill and Melinda Gates Foundation. In an article entitled “How Bill Gates Pulled off the Swift Common Core Revolution,” Layton reveals that the Gates Foundation supplied more than \$200 million not only to the actual development of the CCSS itself but also to building political support across the nation—often convincing state officials to make systematic and expensive changes in their curricular aspirations. Moreover, the foundation spread funds across the entire political spectrum, distributing dollars galore to the two major U.S. teachers unions and such business groups as the U.S. Chamber of Commerce—organizations that have historically clashed, but soon became outspoken proponents of the Common Core. As Layton reports, within two years of the Gates Foundation’s decision to support the Common Core, 45 states and the District of Columbia had fully endorsed the CCSS.

But the curricular aims embodied in the CCSS were destined to serve as much more than lofty statements of curricular intent that, like so many previously crafted sets of curricular aims, typically languished in rarely read reports. This is because, soon after the release of the CCSS in mid-2010, the federal government announced its intention to fund one or more consortia of states whose mission it would be to create assessments suitable for measuring students’ mastery of the skills and knowledge embodied in the CCSS. Two such assessment consortia were selected by federal authorities (from competing bidders) and were funded with approximately \$175 million each to create assessments that, by the 2014–15 school year, could be used to determine students’ mastery of the CCSS. The two consortia were the *Partnership for the Assessment of Readiness for College and Careers (PARCC)* and the *Smarter Balanced Assessment Consortium (SBAC)*. Each of the consortia was initially composed of about 20 to 25 states, all of which agreed to promote students’ mastery of the curricular goals represented by the CCSS.

It should be clear that the nature of the assessments devised by PARCC and SBAC would most likely have a considerable impact on America’s public schools. Because the curricular aims being pursued by so many states would be identical, and the assessments used in those states would also be identical, comparisons among states’ student performances would now be possible in ways that heretofore were impossible. The evaluative impact of such evidence, of course, is apt to be substantial.

As the assessments created by the two consortia became more widely understood, it has become less likely that the sorts of straightforward comparisons among states—comparisons originally foreseen by most proponents of the two assessment consortia—would be less likely to be present. Not only are the reporting categories and the cut-scores set by the two consortia dissimilar, but states are being allowed to infuse unanticipated degrees of local determination into what’s taught and what’s tested. In the middle of 2015, it appeared that considerable uncertainty existed regarding the degree to which identical curricular aims would be pursued by most of the 50 states, and how students’ mastery of those states would be measured.

It is reasonably safe to assume that, under whatever federal revision of ESEA ultimately is enacted by Congress, there will continue to be state accountability tests. The nature and number of those tests may be modified in any ESEA reauthorization, of course, but it seems likely that in one form or another, we will continue to see federal laws calling for state-operated accountability tests. Perhaps those state tests will have been chosen from the CCSS tests provided by one of the two federally funded assessment consortia. Perhaps a state's accountability tests will be state-grown rather than consortium-built. But, one way or the other, state-level accountability tests are apt to be with us for a long while to come. That's the premise that will be employed in the coming pages.

An Updating of the *Standards for Educational and Psychological Testing*

The *Standards for Educational and Psychological Testing* (2014), first published in 1966, contains a set of professionally approved expectations for the way educational and psychological tests ought to be built and used. The *Standards* contain not only a series of comments regarding the way that educational and psychological tests should be evaluated, but they also lay out a specific series of detailed “standards,” that is, mandates regarding what is appropriate in the nature and use of educational and psychological tests. This significant document is published by the American Educational Research Association (AERA) and is approved by that organization as well as the American Psychological Association (APA) and the National Council on Measurement in Education (NCME).

Because the *Standards* are often invoked in high-visibility courtroom contests involving educational tests, their influence on members of the educational measurement community is considerable. Thus, for example, those who write textbooks about educational testing almost always try to make sure what they are recommending in those textbooks is in accord with the latest rendition of the AERA, APA, NCME *Standards*. (I readily count myself among those writers who defer to the *Standards* when recommending how to play in the educational-testing sandbox.)

Periodically revised, for about one and a half decades the 1999 version of the *Standards* held sway, because until mid-2014 the 1999 *Standards* were essentially the only game in town. During that 1999–2014 period, a series of extraordinarily important uses of educational testing took place (for example, the role of students' test scores in educational accountability programs such as those fostered by NCLB). Not surprisingly, then, the 1999 *Standards* were regarded by many educators as being somewhat out of date. And so, when, after a 5-year revision and review process, the 2014 edition of the *Standards* was published, great interest in their contents was predictably displayed by assessment specialists. To illustrate, if pivotal concepts regarding educational assessment had been altered, or even if such concepts had been more clearly explicated, these alterations and these clarified explications would, in a very few years, be incorporated into the set of guide-

lines governing not only what is being professionally recommended about educational testing, but what educators ought to be learning about educational testing.

Candidly, I had been putting the bulk of this eighth edition of *Classroom Assessment* on hold until the updated version of the *Standards* hit the streets. I was reluctant to be advocating practices that might have been acceptable in 1999, but had been meaningfully modified in the new incarnation of the *Standards*. Happily, even though the final publication of the new *Standards* was many months overdue—largely due to the stringent level of scrutiny to which the revised testing standards were subjected by review groups representing AERA, APA, and NCME—the 2014 *Standards* appeared in time to have its contents completely integrated into this edition of *Classroom Assessment*. Although its publication in July 2014 caused the validity and reliability chapters in *Classroom Assessment* to be largely rewritten, at least what you will be reading in the remainder of this book will be in accord with the new *Standards*.

Although I run the risk of oversimplifying a bit, my take on the new *Standards* is that they do not introduce any dramatic reconceptualizations of the fundamental notions of educational testing that have guided educational measurement specialists since the 1999 version of the *Standards*. However, I think the new edition of this potent document both clarifies and tightens the interpretation of several key concepts in educational assessment. We will consider the most salient of those clarified “tightenings” in Chapters 3 and 4 regarding reliability and validity. The 2014 *Standards* did, however, more clearly emphasize the importance of assessment *fairness* than had been seen in earlier revisions. Thus, in the new *Standards* it is appropriate to assert that the three chief emphases are validity, reliability, and fairness.

Do teachers need to become knowledgeable regarding what’s contained in the new 2014 *Standards*? I don’t think so. Let the educational measurement specialists of America fuss with adhering to and interpreting content in the new edition of the *Standards*. But it is a reasonable expectation that teachers at least realize that the ground-rules of educational assessment did not arrive from outer space or from a far Eastern measurement guru. No, these nuts and bolts guidelines about educational testing undergo a rigorous review, rewriting, and approval process every decade or two by three national organizations most concerned with such testing. What teachers need to know, however, is that if they ever find themselves embroiled in any sort of test-related controversy, there exists an authoritative collection of definite dos and don’ts that can be consulted. It is called the *Standards for Educational and Psychological Testing* (2014) and it is available to all.

ASSESSMENT VERSUS TESTING

So far, I’ve been contrasting teaching to testing when, if you’ll glance at this book’s cover, you’ll find that it’s supposed to be a book about assessment. If you’re alert, you’ve already started to wonder—What’s this author trying to pull off? Am I going to learn about *testing* or am I going to learn about *assessment*? Is *assessment*