1 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, backpedalling 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*

simply a more fashionable word for *testing*? In short, what's he up to? These are reasonable questions, and I'll now try to supply you with a set of compelling, confidence-engendering answers.

Almost everyone knows about the kinds of tests typically encountered in school. Most of today's adults, indeed, were on the receiving end of a hoard of teacher-dispensed tests during their own days in school. There were final exams, midterm exams, end-of-unit tests, pop quizzes, and (in the interest of gender equity) mom quizzes. All of those tests had one thing in common. They represented the teacher's attempt to get a fix on how much the teacher's students had learned. More accurately, such tests were employed to determine a student's status with respect to the knowledge or skills the teacher was attempting to promote. This is an altogether praiseworthy endeavor for teachers—to find out how much students know. If teachers are reasonably sure about what their students currently know, then teachers can more accurately tailor any future instructional activities to promote what their students need to learn.

The sorts of tests referred to in the preceding paragraph, such as the quizzes and examinations most of us took in school, have historically been paper-and-pencil instruments. When I was a student, many years ago, the three most common forms of tests I encountered were essay tests, multiple-choice tests, and true-false tests. Until the past decade or so, those three kinds of tests were, by far, the most prevalent sorts of tests found in classrooms.

In recent years, however, educators have been urged to broaden their conception of testing so students' status is determined via a wider variety of measuring devices—a variety extending well beyond traditional paper-and-pencil tests. The reason teachers have been challenged to expand their repertoire of testing techniques is not merely for the sake of variety. Rather, thoughtful educators have recognized there are a number of important kinds of student learning not measured most appropriately by paper-and-pencil tests. If, for example, a teacher wants to determine how well students can function orally in a job-interview situation, it's pretty clear that a written true-false test doesn't cut it.

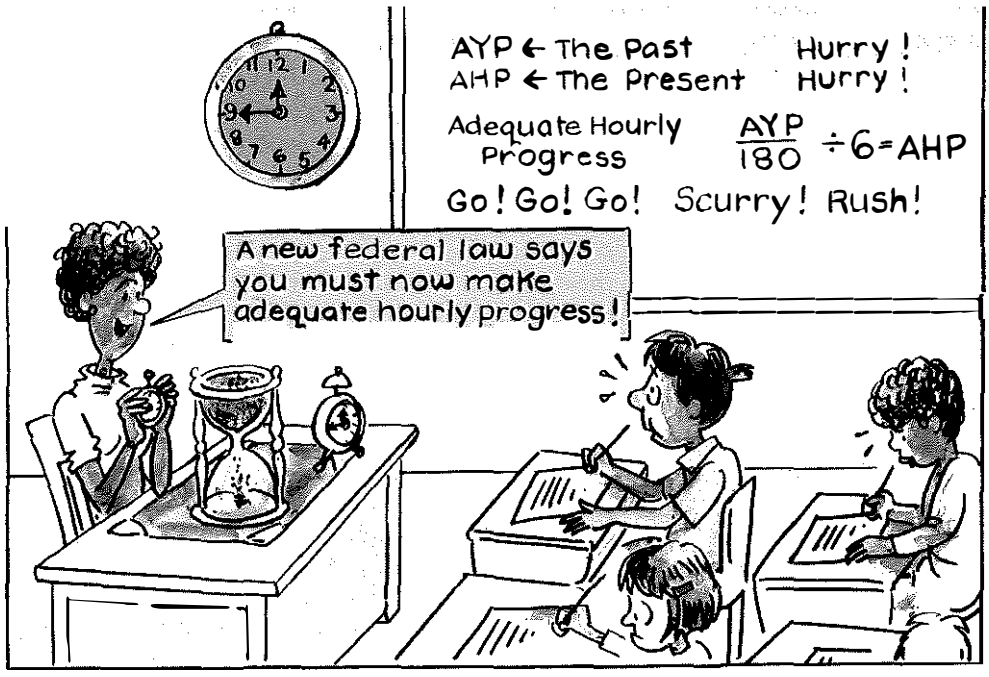
Thus, because there are many worthwhile learning outcomes not best measured by paper-and-pencil tests, and because when most people use the word *test* they automatically think of traditional paper-and-pencil tests, the term *assessment* has been increasingly adopted by many educators and measurement specialists. Assessment is a broader descriptor of the kinds of educational measuring teachers do—a descriptor that, while certainly including traditional paper-and-pencil tests, covers many more kinds of measurement procedures. Here is a working definition of ***assessment*** as it is used in an educational context:

> *Educational assessment is a formal attempt to determine students' status with respect to educational variables of interest.*

Lest you be put off by this fairly foreboding definition, let's briefly consider its chief elements. Note that the kind of assessment we're talking about is aimed

at determining the status of students regarding "educational variables of interest." *Variables* are merely things that vary. (I suspect you could have figured this out all on your own!) In education, for example, we find that students vary in how much they know about a subject, how skilled they are in performing such operations as long division, and how positive their attitudes are toward school. Those are the sorts of variables with which a teacher is typically concerned; thus they are the "variables of interest" teachers typically measure. If the teacher's instructional focus is on the industrial revolution, then the teacher may wish to assess how much students know about the industrial revolution. In that case, the variable of interest would be the degree of students' knowledge regarding the industrial revolution. If the teacher is interested in how confident students are regarding their own written composition skills, then students' composition confidence would be a variable of interest. Educational assessment deals with such variables.

Our working definition also indicates that educational assessment constitutes a "formal" attempt to get a fix on students' status. As human beings, we make all sorts of informal determinations regarding people's status. For example, we may conclude that the woman who cut into the supermarket line ahead of us is rude, or that the man who keeps stumbling as he climbs a set of stairs is clumsy. But these are informal status determinations. Teachers, too, make informal judgments about their students. For instance, a teacher might conclude that a student, based on the student's glum demeanor during the first few moments of class, is definitely



grumpy. Such informal appraisals, although they may be useful to teachers, should not be regarded as educational assessment.

When I was a high school teacher, for example, I employed informal judgment to conclude that Raymond Gonty, one of the seniors in my U.S. Government class, was not really interested in what I was teaching. I reached this conclusion chiefly because Raymond usually slept during class. I became more firmly convinced, however, when he began arriving at class carrying a pillow!

The kind of educational assessment you'll be reading about in this book is formal—that is, it's a deliberate effort to determine a student's status regarding such variables as the student's knowledge, skills, or attitudes. The kind of educational assessment you'll be considering is more than a teacher's "impressions." Rather, you'll be learning about systematic ways to get a fix on a student's status.

Assessment, therefore, is a broad and relatively nonrestrictive label for the kinds of testing and measuring teachers must do. It is a label to help remind educators that the measurement of students' status should include far more than paper-and-pencil instruments. Assessment is a word that embraces diverse kinds of tests and measurements. In the remaining pages, you'll find that although I'll use the term *assessment,* I'll often use the words ***test*** and ***measurement.*** I'll not be trying to make any subtle distinctions at those times. Instead, I'm probably just tired of using the A-word.

WHY SHOULD TEACHERS KNOW ABOUT ASSESSMENT? YESTERYEAR'S ANSWERS

Let's play a bit of time travel. Suppose you were magically transported back to the 1950s or 1960s. And, as long as we're in a let's-pretend mode, imagine you're a new teacher taking part in a fall orientation for first-year teachers in a large school district. The thematic topic of the particular session you're attending is Why Should Teachers Know about Testing? The session's lecturer, Professor Tess Tumm, is supplying the audience with a set of traditional answers to this thematic question based on how teachers actually can use classroom tests. Because you are a docile new teacher (remember, this is imaginary), you are compliantly taking notes to help guide you during the coming school year.

What I'm suggesting, as you've probably guessed, is that there are a number of fairly traditional answers to the question of why teachers should learn about assessment. Those answers have been around for several decades. There is also a set of more current answers to the question of why teachers should know about assessment. Let's give tradition its due and, initially, consider four time-honored answers to the question of why teachers should know about testing. Although these reasons for knowing about classroom assessment may have been around for a while, they're still compelling because they are rooted in the realities of what skilled teachers can do with classroom assessment. These four reasons may well

have been the major points treated by Professor Tumm during our imaginary orientation session of yesteryear.

Determining Students' Current Status

One important reason that a teacher might assess students is to determine what they presently know and can do—for instance, what a group of students' current levels of knowledge are or what their current cognitive skills happen to be. If, for example, a teacher has been instructing students about a series of mathematical operations, there are moments during an instructional sequence when it would be useful for the teacher to know which of those operations have been mastered—and which ones haven't. Based on students' test performances, then, a teacher can decide which mathematical operations seem to need more instructional attention and which ones seem to have been mastered by the students.

There is one oft-encountered instance in which teachers can benefit considerably by using tests to determine students' current status, and it comes up many times during a school year. When teachers are trying to promote their students' attainment of knowledge or skills that are relatively *new* to the students, it is remarkably helpful to get a fix on what it is that students already know and can do.

For instance, if Jaime is already truly proficient in solving simultaneous equations, it's a waste of Jaime's time to make him plow through practice piles of such equations. When I was growing up, the expression "That's like carrying coal to Newcastle" was used to disparage any scheme that reeked of redundancy. (I always assumed that coal mining was a big deal in Newcastle.) Well, teachers who relentlessly keep instructing students regarding knowledge or skills that the students have already mastered are definitely lugging coal lumps to Newcastle. Assessment can allow teachers to identify students' current capabilities and, as a consequence, can help teachers avoid superfluous and wasteful instruction.

Thus, by measuring students' status, teachers can discern (1) where to put their instructional energies to ameliorate a student's shortcomings and (2) what already mastered skills or knowledge can be instructionally avoided. Such assessment is particularly useful for a teacher's planning if the assessment is carried out at the beginning of an instructional sequence. This kind of early diagnosis is often referred to as ***preassessment*** because it is assessment that takes place prior to the teacher's initiation of instruction.

Monitoring Students' Progress

A second, related answer to the question, Why should teachers assess? is that such assessments help teachers determine whether their students are making satisfactory progress. Sometimes, of course, it's easy for teachers to tell whether their students are or are not progressing satisfactorily. I can still recall, with suitable embarrassment, the absolutely scintillating lesson I provided as a high school English teacher on the topic of Modifying Gerunds with Possessives. It was a lesson designed for

a full-class period, and I was confident that at its conclusion my students would not only understand the topic but also be able to explain to others why one of the following sentences contains an appropriate pronoun and one does not:

Improper Pronoun ↓ *Gerund* ↓

☹ **Sentence 1:** I really appreciate you sending the brownies.

Proper Pronoun ↓ *Gerund* ↓

☺ **Sentence 2:** I really appreciate your sending the brownies.

At the end of a bravura 40-minute lesson, replete with all sorts of real-life examples and a host of on-target practice activities, I was certain I had effectively taught students that because a gerund was a noun-form of a verb, any modifiers of such gerunds, including pronouns, must be possessive. And yet, at the end of the lesson, when I looked into my students' baffled faces, I realized that my optimism was unwarranted. After asking several students to explain to me the essence of what I'd been talking about, I quickly discerned that my lesson about gerund modifiers was not an award-winning effort. Most of my students couldn't distinguish between a gerund and a geranium.

Although teachers can occasionally discern informally, as I did, that their students aren't making satisfactory progress, more often than not we find teachers' believing their students are progressing quite well. (Note in the previous sentence that the modifier of the gerund *believing* is the possessive form of *teachers*. Yes, I'm still trying.) It's only human nature for teachers to believe they're teaching well and their students are learning well. But unless teachers systematically monitor students' progress via some type of assessment, there's too much chance that teachers will improperly conclude progress is taking place when, in fact, such progress is not.

A useful function of classroom assessment, therefore, is to determine whether students are moving satisfactorily toward the instructional outcomes the teacher is seeking to promote. If progress for all students is satisfactory, of course, then the teacher need make no instructional adjustments. If progress for most students is satisfactory, but a few students are falling behind, then some separate doses of remedial assistance would seem to be in order. If progress for most students is inadequate, then the teacher should substantially modify whatever instructional approach is being used because, it is all too clear, this approach is not working. Progress monitoring is a time-honored and altogether sensible use of classroom assessment.

I've run into a number of teachers in recent years who refer to this use of assessment as "dip-sticking." When I think back to the days that I occasionally used a dip-stick to determine if my second-hand automobile was running low on oil (as it typically was), that label definitely rings true.

A teacher ought to monitor students' progress via classroom assessment because, more often than you'd think, the teacher can stop instructing on a certain topic well in

advance of what the teacher had anticipated. Suppose, for instance, you're attempting to get your students to acquire a certain skill, and you've set aside two weeks to promote their mastery of the skill. If you monitor students' progress with an assessment after only a week, however, and discover your students have already mastered the skill, you should simply scrap your wee k-two plans and smilingly move on to the next topic.

Another way of thinking about the monitoring of student progress is that it positions teachers to use the results of classroom tests as part of *formative assessment*—that is, the use of assessment-elicited evidence intended to improve unsuccessful yet still modifiable instruction. *Summative assessment,* in contrast, refers to the use of tests whose purpose is to make a final success/failure decision about a relatively unmodifiable set of instructional activities. In a review of research studies focused on the instructional payoffs of formatively oriented classroom assessment, two British investigators (Black and Wiliam, 1998) concluded that the use of progress-monitoring classroom assessments can promote striking gains in student learning on both teacher-made and external exams.

Based on the Black and Wiliam conclusions regarding the major instructional dividends of formatively oriented classroom assessment, members of Britain's Assessment Reform Group introduced the idea of classroom assessment *for* learning—in contrast to assessment *of* learning. They describe this approach as follows:

> *Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purpose of accountability, or of ranking, or of certifying competence. (Black et al., 2002)*

Stiggins and Chappuis (2012) have also pushed assessment for learning as the cornerstone of effective classroom measurement. Later, in Chapter 12, you will learn much more about the fundamentals of formative assessment.

Assigning Grades

If I were somehow able to carry out an instant nationwide survey of beginning teachers and asked them, "What is the most important function of classroom assessment?" I know what answer I'd get from most of the surveyed teachers. They'd immediately respond: *to give grades.*

That's certainly what I thought testing was all about when I taught in public schools. To be honest (confession, I am told, is good for the soul), the *only* reason I tested my students was to give them grades. I've talked to hundreds of teachers during the past few years, and I've been dismayed at how many of them continue to regard testing's *exclusive* function to be grade giving. A third reason, therefore, that teachers assess students is to assemble the evidence necessary to give their students grades. Most school systems are structured so the end-of-course or end-of-year grades a student earns constitute the beginnings of a record of the student's personal accomplishments—a record destined to follow the student throughout life. Thus, it is imperative teachers not assign grades capriciously. Whether we like it or not, students' grades are important.

The best way to assign grades properly is to collect evidence of a student's accomplishments so the teacher will have access to ample information before deciding whether to dish out an A, B, C, D, or F to a student. Some school systems employ less traditional student grading systems—for example, the use of descriptive verbal reports that are relayed to parents. Yet, whatever the reporting system used, it is clear the teacher's assessment activities can provide the evidence necessary to make sensible student-by-student appraisals. The more frequent and varied the evidence of student accomplishments, the more judiciously the teacher can assign to students the grades they deserve.

A corollary principle linked to "tests as grade determiners" is that some teachers also employ the prospect of upcoming tests to motivate their students. Because a student's grade is often dependent on the student's test performances, teachers will frequently employ admonitions such as, "Be sure to study this chapter carefully, because you have an important end-of-chapter exam coming up on Thursday!" Some teachers surely employ impending tests as a motivational device.

In recent years, several thoughtful educators have proffered sensible guidance regarding how teachers ought to award grades to their students (for example, Guskey, 2015). A consensus of these writers' thinking—a consensus focused on "standards-based" grading—will be presented in Chapter 16 to wrap up this edition of *Classroom Assessment.*

Determining One's Own Instructional Effectiveness

A fourth and final reason teachers have traditionally been told they should test students is that students' test performances can help teachers infer how effective their teaching has been. Suppose a teacher sets out to have students master a set of worthwhile skills and knowledge regarding Topic *X* during a 3-week instructional unit. Prior to instruction, a brief test indicated students knew almost nothing about Topic *X* but, after the unit was concluded, a more lengthy test revealed students had mastered most of the skills and knowledge addressed during the Topic *X* unit.

Because the comparison of students' pretest and posttest results indicated the teacher's students had acquired ample knowledge and skills regarding Topic *X*, the teacher has a charming chunk of evidence that the instructional approach being used appears to be working. If the teacher's instruction seems to be promoting the desired outcomes, then it probably shouldn't be altered much.

On the other hand, let's say a teacher's Topic *X* pretest-to-posttest results for students suggest students' progress has been piffling. After comparing results on the end-of-instruction posttest to students' performance on the preinstruction test, it appears students barely knew more than they knew before the instruction commenced. Such trivial student growth should suggest to the teacher that adjustments in the instructional activities seem warranted when teaching Topic *X* again next term or next year.

I'm not suggesting students' pretest-to-posttest results are the only way for teachers to tell whether they're flying or flopping, but students' end-of-instruction performances on assessment devices constitute a particularly compelling indication of whether teachers should retain, alter, or jettison their current instructional procedures.

In review, then, we've considered four fairly traditional answers to the question of why teachers should assess students. Here they are again:

Traditional Reasons That Teachers Assess Students

- To determine students' current status
- To monitor students' progress
- To assign grades to students
- To determine instructional effectiveness

You will notice that each of these four uses of educational assessment is directly related to *helping the teacher make a decision.* When a teacher assesses to determine students' current status, the teacher uses test results to *decide* what instructional objectives to pursue. When a teacher assesses students' progress, the teacher uses test results to *decide* whether certain parts of the ongoing instructional program need to be altered. When a teacher assesses students to help assign grades, the teacher uses students' performances to *decide* which students get which grades. And, finally, when a teacher uses pretest-to-posttest assessment results to indicate how effective an instructional sequence has been, the teacher is trying to *decide* whether the instructional sequence needs to be overhauled. Teachers should never assess students without a clear understanding of what the decision is that will be informed by results of the assessment. The chief function of educational assessment, you see, is to improve the quality of educational decision-making.

Taken in concert, the four traditional reasons just described should incline teachers to assess up a storm in their classrooms. But these days even more reasons can be given regarding why teachers need to know about assessment.

WHY SHOULD TEACHERS KNOW ABOUT ASSESSMENT? TODAY'S ANSWERS

In addition to the four traditional reasons teachers need to know about assessment, there are three new reasons that should incline teachers to dive joyfully into the assessment pool. These three reasons, having emerged during the past decade or so, provide compelling support for why today's teachers dare not be ignorant regarding educational assessment. Let's consider three new roles for educational assessment and see why these new functions of educational testing should incline you to feverishly pump up your assessment knowledge and skills.

Influencing Public Perceptions of Educational Effectiveness

When I was a high school teacher a long while ago, teachers were occasionally asked to give nationally standardized achievement tests. But, to be honest, no one really paid much attention to the test results. My fellow teachers glanced at the test-score reports, but were rarely influenced by them. The public was essentially

oblivious of the testing process and altogether disinterested in the results unless, of course, parents received a report that their child was performing below expectations. Testing took place in the fifties and sixties, but it was definitely no big deal.

During the seventies and eighties, however, a modest journalistic wrinkle changed all that. Newspaper editors began to publish statewide educational test results on a district-by-district and even school-by-school basis. Citizens could see how their school or district stacked up in comparison to other schools or districts in the state. Districts and schools were *ranked* from top to bottom.

From a news perspective, the publishing of test results was a genuine coup. The test scores were inexpensive to obtain, and readers were really interested in the scores. Residents of low-ranked districts could complain; residents of high-ranked districts could crow. More importantly, because there are no other handy indices of educational effectiveness around, test results became the measuring-stick by which citizens reached conclusions about how well their schools were doing. There are many reports of realtors trying to peddle homes to prospective buyers on the basis that a house was located "in a school district with excellent test scores."

Let me be as clear as I can possibly be about this issue, because I think it is a terrifically important one. As matters stand, students' performances on a state's accountability tests are certain to influence the way that *all* teachers are evaluated—even if a particular teacher's own students never come within miles of an accountability test. Here's how that will happen.

Suppose you teach ninth-grade social studies, and your ninth-graders aren't required to take federally required accountability tests. Suppose you're a second-grade teacher, and your students aren't required to take any kind of accountability test. Suppose you're a high school teacher who teaches subjects and grade levels where no federal or state accountability tests are required. In all these "suppose" situations, *your* students won't be taking accountability exams. However, the public's perception of *your* personal effectiveness will most certainly be influenced by the scores of *your* school's students on any accountability tests that are required for such schools. Let's be honest—do you want to be a teacher in a "failing" school? Do you want your students' parents to regard *you* as ineffective because you happen to do your teaching in what's thought to be a sub-par school? I doubt it.

The reality is that the performance of any school's students on federally stipulated accountability tests will splash over on every teacher in that school. If you teach in a school that's regarded as successful, then you will be seen as a member of an effective educational team. The opposite is also true. Unless federal accountability requirements are substantially softened, no public school teacher will be able to remain isolated from the impact of externally imposed accountability tests.

And, as I'll try to point out later in the book, the nature of a school's success on high-stakes external assessments, such as federally required accountability tests, will (and, indeed, *should*) have an impact on the sorts of classroom assessments you personally choose to employ. We live in an era when public perceptions of schooling are more important than some educators might prefer. Yet, like it or not, that's the reality today's teachers must face.

DECISION TIME

Pressure from "Higher Ups"

Laura Lund has been teaching second-graders at Horace Mann Elementary School for the past 3 years. During that period, Laura has become increasingly convinced that "developmentally appropriate instruction" is what she wants in her classroom. Developmentally appropriate instruction takes place when the instructional activities for children are not only matched with the typical developmental level of children in that grade but also matched with the particular developmental level of each child. Because of her growing commitment to developmental appropriateness, and its clear implications for individualized instruction, Laura's students now no longer receive, in unison, the same kinds of massed practice drills in reading and mathematics Laura provided earlier in her career.

Having discovered what kinds of changes are taking place in Laura's second grade, however, the third-grade and fourth-grade teachers in her school have registered great concern over what they regard as less attention to academics, at least less attention of the traditional sort. Because state accountability tests are given to all third- and fourth-grade students each spring, Laura's colleagues are afraid their students will not perform well on those tests because they will not be skilled at the end of the second grade.

A year or so earlier, when Laura was teaching her second grade in a fairly traditional manner, it was widely recognized that most of her students went on to the third grade with a solid mastery of reading and mathematics. Now, however, the school's third- and fourth-grade teachers fear that "Horace Mann's accountability scores may plummet."

As Laura sees it, she has to decide whether to (1) revert to her former instructional practices or (2) maintain her stress on developmentally appropriate instruction. In either case, she realizes that she has to try to justify her action to her colleagues.

If you were Laura Lund, what would your decision be?

Helping Evaluate Teachers

Teaching skill is coming under increasing scrutiny these days. With the push for more rigorous evaluation of a classroom teacher's performance, we now see many teacher appraisal systems in which students' test performances constitute one key category of evidence being used to evaluate teachers. Sometimes, teachers are directed to assemble pretest and posttest data that can be used to infer how much learning by students was promoted by the teacher. And, of course, teachers whose students are *required* to take a state's annual accountability tests understand all too well that their students' scores on those tests will play a prominent role in teacher evaluation—that is, in the evaluation of *their* teaching.

Although we will consider the topic of teacher evaluation far more thoroughly in Chapter 15, it should be noted at this point that a pair of federal initiatives have spurred much greater use of students' test scores in the appraisal of teachers. In 2009, the federal

Race to the Top Program offered some serious financial grants to states that would be willing, among other reforms, to install teacher evaluation systems in which students' test performances played a prominent role. Two years later, in 2011, once again federal officials offered the ESEA Flexibility Program that allowed states to seek a waiver from the harsh penalties linked to the final days of the No Child Left Behind Act. In that second initiative, states were once more informed that they had a better chance of snaring a waiver from Washington, DC, if they installed teacher evaluation programs in which students' test scores were regarded as a significant factor in evaluating the state's teachers.

Even though the education officials of most states sought one or both of these federal incentive programs, and promised to implement systems for evaluating teachers (and principals) using programs featuring student's assessed growth, a good many states now seem to be treading water regarding the implementation of their promised educator evaluation programs. Nonetheless, in all but a few of our states, descriptions of the current state-decreed teacher evaluation system calls for use of students' *measured* growth as one key evaluative criterion.

As a practical matter, then, because educational assessments will be employed to collect evidence of students' learning, and because this evidence will be used to evaluate teachers, a teacher would have to be a downright dunce to dodge the acquisition of information about sensible and senseless ways to measure students' status.

However, as you will learn in later chapters, only certain kinds of educational assessments can properly carry out this sort of test-rooted task. Most of the tests proposed for this purpose are altogether inappropriate for such an evaluative assignment. Nonetheless, if judgments about teachers' quality are—because of well-intentioned legislative actions—to be based in part on students' assessment performances, then it is apparent that teachers need to learn about the kinds of tests that will support or, possibly, distort this sort of evaluative endeavor.

Experienced teachers will be quick to tell you that the caliber of students' test performances is dramatically influenced by the caliber of the students being tested. It should be apparent that a teacher who is blessed with a flock of bright students will almost always get better test results than a teacher who must work with a less able group of students. And let's not forget about the quality of students' previous teachers. Wouldn't you rather be receiving a new group of students who had been effectively taught by Mrs. X than a group of students who had been ineffectively taught by Mrs. Y? Nonetheless, increasing numbers of statewide and districtwide teacher evaluation systems now call for teachers to assemble tangible evidence of student accomplishments based on external exams or teacher-made classroom assessments. It is clear, therefore, that today's teachers need to know enough about educational assessment so they can corral compelling evidence regarding their own students' growth. We will consider today's teacher evaluation tempest in more detail in Chapter 15.

Clarifying Teachers' Instructional Intentions

For many years, educational tests were regarded as instructional afterthoughts. As soon as an instructional unit was over, the teacher got busy cranking out a test. Tests were rarely created before instruction was initiated. Instead, tests were devised *after* instruction to fulfill some of the traditional functions of

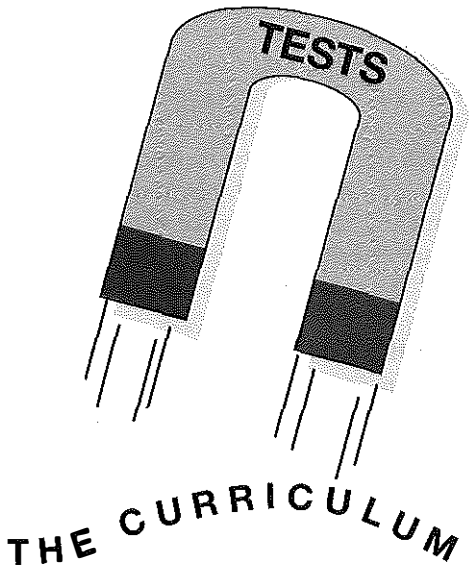
educational assessment described earlier in the chapter—for example, the assignment of grades.

Today, however, many educational measuring instruments have become high-stakes tests. A ***high-stakes test*** is an assessment for which important consequences ride on the test's results. One example of an educational high-stakes test would be a statewide basic skills test that must be mastered before a student graduates. (Note that the important consequences are for the test taker.) Another example would be the results of a districtwide achievement test that are publicized so local taxpayers' judgments about educational effectiveness are influenced by the test results. (Note that in this second case the important consequences apply to the educators who prepared the students, not the test-takers themselves.)

A federally required accountability test will fall into this second category of high-stakes tests. Because students' performances on these tests will so powerfully influence the way people regard a school staff's quality, such accountability tests will be genuinely high stakes in nature. You should know, however, there was nothing in NCLB that required diploma denial or that obliged students to be held back at grade level if they fail to perform well enough on a test. A *state's* decision can transform a federal test into one that has an adverse impact on a particular student. Many people continue to be confused by this, for they assume that any federally mandated accountability test automatically requires diploma-denial or promotion-denial testing. It's just not so.

Insofar as important consequences are directly linked to assessment results, the content of such high-stakes tests tends to be emphasized instructionally by teachers. Because teachers want their students to perform well on high-stakes tests (for the students' own good and/or for the teacher's benefit), high-stakes tests tend to serve as the kind of curricular magnet seen in Figure 1.1.

figure 1.1 ▪ The Curricular Impact of High-Stakes Tests



On some educational grounds, teachers might prefer that tests did not influence instruction so directly, but the reality is that high-stakes assessment will definitely have an impact on classroom instructional practices. Because this curricular influence is certain to be present, it will be in teachers' and students' best interests if the nature of the upcoming assessment is sufficiently well understood so the teacher can organize the most effective, on-target instruction possible. (Later in the book, we will consider the deficits of teaching exclusively toward assessment targets.) In a sense, however, the more that teachers understand what the innards of a test are, the more effectively they can use this understanding to clarify what's to be sought instructionally.

Earlier in the chapter, the *Common Core State Standards* were identified as a set of widely adopted curricular goals for many of our states. It was also pointed out that two assessment consortia (SBAC and PARCC) had been commissioned with substantial funds to create assessments intended to measure students' mastery of the curricular aims embodied in the *CCSS*. Well, when many of the nation's educators attempted to promote students' attainment of those curricular aims, they soon realized that although teachers could get a rough, general idea of what learner outcomes were sought by the *CCSS,* it was really necessary to await the release of the PARCC and SBAC test items before one could really know what was meant by many of the *CCSS* goals. Curricular aims are often just words—words so ambiguous in their meaning that it is sometimes necessary to use tests to operationalize those nice-sounding but often too general words. Thus, when SBAC and PARCC released many sample items in 2014 and 2015, American educators arrived at a much clearer idea of what the *CCSS* curricular aims actually mean.

Even the low-stakes classroom tests routinely employed by teachers can be used to help teachers clarify their instructional targets. Tests should obviously not, then, be instructional afterthoughts. Rather, classroom assessment instruments should always be prepared *prior* to any instructional planning in order for the teacher to better understand what is being sought of students and, therefore, what to incorporate in instructional activities for students. Assessment instruments prepared prior to instruction concretely exemplify a teacher's instructional intentions and, as a consequence, clarify those intentions. Clarified instructional intentions characteristically lead to more effective instructional decisions by the teacher. The better you understand where you're going, the more efficiently you can get there.

To reiterate, we've now looked at three reasons today's teachers, unlike their counterparts of a few decades ago, need to know about assessment. These reasons are supplemental to, not in place of, the previously considered traditional reasons teachers assess students. Here are the three new reasons for teachers' familiarization with educational assessment:

Today's Reasons for Teachers to Know about Assessment

- Test results determine public perceptions of educational effectiveness.
- Students' assessment performances are increasingly being included as part of the teacher evaluation process.
- As clarifiers of instructional intentions, assessment devices can improve instructional quality.

These reasons are also linked to decisions. For instance, when citizens use test results to reach judgments about a school district's effectiveness, those judgments can play a major role in determining what level of taxpayer support will be provided in that district. There are also decisions on the line when students' test scores are used as evidence to evaluate teachers. Such decisions as whether the teacher should be granted tenure or receive merit-pay awards are illustrative of the kinds of decisions that can ride, at least in part, on the results of educational assessments. Finally, from the teacher's perspective, when tests serve as clarifiers of the teacher's instructional intentions, the teacher can make better decisions about how to put together instructional activities likely to help students attain the instructional intentions represented by the assessment. With these three current roles of educational assessment, as was true with the four more traditional roles of educational assessment, test results should contribute to educational *decisions*.

WHAT DO CLASSROOM TEACHERS REALLY NEED TO KNOW ABOUT ASSESSMENT?

Whether you are already a teacher or are preparing to become a teacher, you really do need to know about educational assessment. But the field of educational assessment contains huge chunks of information. In fact, some educators devote their entire careers to assessment. Clearly, there's more to educational assessment than you probably care to know. The question is, What should *classroom teachers* know about assessment?

The title of this book suggests an answer—namely, *Classroom Assessment: What Teachers Need to Know.* The key word in the title, at least for purposes of this discussion, is *need.* There are oodles of fascinating things about assessment you might learn. You'd even find a few of them interesting (not all that many, I suspect). But to help your students learn, you really don't *need* to know a host of assessment esoterica. This book about educational assessment is deliberately focused on only those things that you really *must know* in order to promote your students' learning most effectively. I am altogether reluctant to clutter your skull with a galaxy of nice-to-know but nonessential knowledge about educational assessment. Such nice-to-know content often crowds out the need-to-know content. There is, after all, only so much skull space available.

As a preview, I want to describe briefly what you will have learned by the time you reach the book's index. (I have never had much confidence in any book's index as a teaching tool, so if you haven't learned what's needed by that time—it's likely to be too late.) It may be easier for you to get a handle on what you'll be reading if you realize you'll be covering topics dealing chiefly with:

1. Constructing your own assessment instruments
2. Using assessment instruments constructed by others
3. Planning instruction based on instructionally illuminating assessments

Creating Classroom Assessment Devices

Let's start with the kinds of classroom assessment devices you will personally need to create. The chief thing you will learn in this book is how to construct a wide variety of assessment instruments you can use as part of your day-to-day classroom instruction. You really do need to know how to determine what your students have learned—for example, whether they comprehend what they have read. You also really do need to know how to get a fix on your students' educationally relevant attitudes—such as how positively disposed your students are toward the subject(s) you're teaching. Thus, you are going to be learning about how to create classroom assessment approaches to measure students' *achievement* (that is, the knowledge and/or skills students acquire) as well as students' *affect* (that is, the educationally pertinent attitudes, interests, and values influenced by school).

As suggested earlier, the kinds of classroom assessment procedures you'll be learning about will extend well beyond traditional paper-and-pencil testing instruments. You may even learn about several assessment approaches with which you are currently unfamiliar.

In a related vein, you will also learn how to judge the quality of the assessment devices *you* create. And, at the same time, you will learn how to judge the quality of assessment devices created by others. Those "others" might be your own colleagues or, perhaps, the folks who devise large-scale assessment instruments such as districtwide or statewide tests. It may seem presumptuous for me to suggest that you, a classroom teacher (in practice or in preparation), could be judging the efforts of folks who create large-scale standardized educational tests. But you'll discover from this book that you will, indeed, possess the knowledge and skills necessary to distinguish between tawdry and terrific practices by those who create such large-scale educational tests. In particular, you can use your new knowledge to judge the quality of any accountability tests that may be used in your state. Those tests are going to be so important that, if you are teaching in a state whose accountability tests are educationally unsound, you definitely need to know it. What can you do if you discover that your state's high-stakes tests are inappropriate? Well, two action-options come quickly to mind. For openers, you can learn enough about the shortcomings of the state's tests so that you are able to explain coherently to your students' parents why the high-stakes tests your state has chosen are unsound. Second, and this may require a nontrivial expenditure of effort on your part, you can take part in educator organizations that are willing to bring about the installation of more suitable educational assessments. You can do neither of these things, of course, if you have only a skimpy knowledge of what makes your state's tests tick—and how they really ought to be ticking.

It is also important for you to know enough about educational assessment so that you can assist your colleagues in evaluating an ever-increasing array of commercially developed educational tests. Educators have seen a spate of such tests developed in the years following emergence of the *Common Core State Standards*. Some of these vendor-produced tests may be quite useful to teachers. Some of those tests, however, are seriously flawed—apparently cranked out merely to bring in a few bucks from desperate educators. There is no guarantee that a published

Assessment Literacy: Only for Grown-ups?

Students' performances on tests can have an enormous impact not only on the students who actually take educational tests but also on the test-taker's family. Teachers, too, are often affected by their students' test scores. To illustrate, today's teachers seem to be frequently buffeted by educational accountability tests on which students' test scores can have an impact on teachers' tenure, assignment, and salaries. Clearly, educators at all levels, whether teachers or administrators, need to learn enough about educational tests to carry out their responsibilities successfully.

What those educators need, then, is a reasonable dose of ***assessment literacy***. And here's a definition of it:

> *Assessment literacy consists of an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions.*

Notice that this definition is focused on someone's understandings of educational measurement's *basic* concepts and procedures of educational assessment. The stuff an assessment-literate person needs to understand is not esoteric and incomprehensible. Rather, most of what an assessment-literate person should know is just common sense applied to educational measurement. Describing assessment literacy a bit differently, it represents the main-line measurement procedures and concepts thought to make a difference in the decisions made about the students who take educational tests.

Well, if that's what assessment literacy is, who needs to have it? There's considerable pressure these days on teachers to become more assessment literate. You are currently reading a book that, unless the book's author has really mucked up, ought to help you personally become more assessment literate. But what about educational policymakers? And what about parents of school-age children? And, finally, ***what about students themselves?*** Don't all three of these groups need to beef up their understandings about the key assessment-related principles and processes that can influence students' lives?

I certainly think so. I hope, as you read this book, you'll occasionally pause to think how you might relay to policymakers (such as school board members), parents (such as (such as your own students' parents) and students (such as the ones you're teaching) the most essential things about the assessment-related concepts and procedures you're encountering in these pages. Remember, because test results these days can increasingly enhance or impair the decisions made about students, don't those students *at least* have the right to know what's going on behind the assessment curtain? A reasonable lump of assessment literacy is good for almost everyone!

test ought ever to have been published. Only an educator who possesses at least a small sack-full of assessment sophistication will be able to tell whether a commercially created educational test is yummy or gummy.

Fundamentally, educational assessment rests on a foundation of common sense. Once you learn the technical vocabulary of assessment, you'll be able to identify departures from commonsensical assessment practices, whether those departures are seen in your own tests, in the tests of a teacher down the hall, or in the tests created by district, state, or national assessment specialists. In short, after you finish this book, you really need not be deferent to any "measurement experts." You'll know enough to spot serious shortcomings in their work.

In Chapter 3, for example, you will learn about three criteria you can use to evaluate all educational tests. Those criteria apply with equal force to tests you might develop as well as to tests that are commercially developed. Once you get the hang of how to evaluate tests, you can apply this evaluative skill in many settings.

Interpreting Standardized Test Results

Because your students will often be assessed with nationally standardized or state-developed tests, you will need to know how to interpret the results of such tests. In general, commercially developed educational tests focus either on students' *achievement,* which, as noted earlier, deals with the knowledge and skills that students have acquired, or commercial tests focus on *aptitude,* which is a term used to describe a student's learning *potential.* You should know, however, that the term *aptitude* is definitely falling from grace these days. In the old days, when I was a tyke, people talked about intelligence. As a prospective teacher, I learned all about the intelligence quotient (IQ), which was a numerical way of indicating the degree to which a particular individual's intellectual abilities exceeded or fell short of conventional expectations for such individuals.

To calculate someone's IQ, you simply divided a student's *mental age* (based on how well a student's test score stacked up against a norm group's scores) by a student's *chronological age* (based on a calendar). The result of this was

$$\frac{\text{MA}}{\text{CA}} = \text{IQ}$$

But "intelligence" has fallen decisively out of favor with educators during the past few decades. The term *intelligence* conveys the notion that students possess an inborn potential about which schools can do little to influence. Yet, the so-called intelligence tests, widely used until recently, often measured what students had learned at school or, more importantly, what students had learned at home. Thus, the term *aptitude* has been increasingly used rather than *intelligence* in order to convey a notion of a student's academic potential. But even the term *aptitude* tends to create the perception there is some sort of innate cap on one's potential. Because of this perception, we now find that the commercial test makers who formerly created so-called intelligence tests, and then renamed them aptitude tests, are looking for a less negatively loaded descriptor. Interestingly, the tests themselves, although they've been relabeled, haven't really changed all that much.

At any rate, you'll learn how to make sense out of the kinds of reports regarding student performance released by those who conduct large-scale assessments. You will

But What Does This Have to Do with Teaching?

This chapter contains over a half-dozen reasons teachers need to learn about assessment. Actually, there are seven reasons in the chapter, and that's one more reason than a half-dozen. (Notice how low-level the arithmetic in this book is going to be!)

But let me single out the *two* reasons I think, *from an instructional perspective,* all teachers need to know about testing. The first of these reasons is the last of the seven reasons cited in the chapter—namely, the *instructional-planning payoffs* teachers can get from a more clear understanding of what they're trying to have their students accomplish. Because a properly constructed classroom test can truly exemplify what a teacher is trying to achieve, the resulting *clarity of intention* helps teachers make more astute decisions when they plan their instruction.

When I recall my own early years as a high school teacher, I remember how often I simply whipped together lesson plans that *seemed* somewhat instructional. Yet, my planning was almost never guided by a truly clearheaded notion of what knowledge, skills, or attitudes I wanted my students to possess when the instruction was over. If I had relied on my classroom tests to clarify what I wanted my students to become, I'm certain my lessons would have been far better focused on my intended instructional outcomes.

The second reason I think all teachers should become more astute regarding assessment is also instructionally rooted. It's the second of the four traditional reasons considered in the chapter—namely, so teachers can *monitor students' progress*. If teachers use students' assessed levels of achievement to determine whether the current instructional plan is stellar or sickly, then teachers' adjustments in lessons can, if warranted, be made. Without the evidence yielded by the classroom formative-assessment process, a teacher will often fail to spot instructional inadequacies. As Black and Wiliam, the British investigators, made quite clear, the instructional dividends from monitoring students' progress can be striking. And their views were based on solid research investigations, not wishful yearnings. My grandparents came to the United States from Great Britain, so I typically groove on anything that's asserted with a British accent. But, then, doesn't everyone?

need this knowledge not only to inform your own decisions about classroom instruction but also to interpret students' test performances to parents who may demand answers to questions such as, "What does my son's standardized test performance at 40th percentile really mean?" or "If my fifth-grade daughter earned a grade-equivalent score at the eighth-grade level on this year's standardized achievement test, why shouldn't she be promoted?" In short, you'll learn how to interpret students' performances on both ***achievement tests*** and ***aptitude tests***. Moreover, given the relatively recent arrival of computer-administered and computer-adaptive educational tests, you'll find that parents are apt to be raising questions about such technologically abetted tests. Teachers need to be able to answer such questions—preferably with the correct answers.

Instructionally Illuminating Assessment

Earlier, it was suggested that because assessment devices exemplify a teacher's instructional intentions, those assessment instruments can clarify the teacher's instructional decision-making. You'll learn more about how the link between testing and teaching can prove beneficial to your students because you can provide more on-target and effective instruction.

On the other hand, you'll also learn how some teachers inappropriately prepare their students for tests, particularly for high-stakes tests. You will learn about ways of judging whether a given test-preparation practice is (1) in students' best interests from an educational perspective and (2) in educators' best interests from an ethical perspective. In short, you'll learn about the increasingly important relationship between instruction and assessment.

Later, you'll learn about a way to build classroom assessments so that they'll have a decisively positive impact on how well you teach. *Tests,* if deliberately created with instruction in mind, *can boost your personal success as a teacher.* We'll dig into that topic in Chapter 12.

There's one other issue that I'd like to bring to your attention—namely, the possibility (after you've finished the book) of your helping *parents* learn more about educational assessment. And why, you might ask, should a teacher be messing around trying to promote parental measurement moxie? It's a good question. And the answer is this: Parents who are *assessment literate* will be better able to *help you help their children learn* more successfully.

You see, most parents know little more about testing than what they can recall, often vaguely, from their own classroom days as students. But the nature of classroom testing has changed dramatically since that time. Not only are there new approaches to assessment being used in classrooms (all of which you'll learn about in this book), but students' test scores are also being used to judge the success of teachers' instructional efforts. You'll learn in Chapter 15 that, depending on the tests being used, this may be a dumb idea. If you and your students' parents truly understand the fundamentals of educational assessment, you can work together in many ways that will benefit your students. Assessment-literate parents can be a potent force to counter the serious misuses of educational tests we see so often today. And, if you are teaching in a state whose officials have opted to use instructionally inappropriate accountability tests, you'll find that assessment-literate parents can be a potent political force who might, if you're lucky, help get more appropriate accountability tests installed in your state.

There's another audience for assessment literacy that you'll hopefully give some thought to as you wend your way through this book. Please recognize that the lives of today's students are increasingly influenced by their performances on various kinds of educational tests. Why not, therefore, provide at least a dose of assessment literacy to students themselves? As you will see, most of the assessment concepts treated in the book are not particularly complicated. Indeed, the truly essential assessment understandings needed by students are well within the grasp of those students. Why not splash a bit of assessment literacy on those students? They really need it!

PARENT TALK

Mr. and Mrs. Smothers are attending a back-to-school night at a middle school where their daughter, Cathy, is a fifth-grader. After briefly leafing through Cathy's math portfolio and language arts portfolio, they get around to the real reason they've come to school. Mrs. Smothers, looking more than a little belligerent, says, "Cathy tells us she gets several teacher-made tests in class every week. All that testing can't be necessary. It obviously takes away the time you spend teaching her! Why is there so darn much testing in your class?"

If I were you, here's how I'd respond to Mrs. Smothers:

"I suppose that it might seem to you there's too much testing going on in my class, and I can understand your concern about testing-time taking away from teaching-time. But let me explain how the time my students spend doing classroom assessments really leads to much *better* use of instructional time.

"You see, the way I use classroom assessment is to make sure my instruction is on target and, most importantly, that I don't waste the children's time. Last month, for instance, we started a new unit in social studies and I gave students a short pretest to find out what they already knew. To my delight, I discovered that almost all of the students—including Cathy—knew well over half of what I had been planning to teach.

"Based on the pretest's results, I was able to shorten the social studies unit substantially, and spend the extra time giving students more practice on their map-interpretation skills. You probably saw some of the maps Cathy was interpreting as part of her homework assignments.

"Mr. and Mrs. Smothers, I want Cathy's time in class to be as well spent as possible. And to make sure of that, I use formal and informal classroom tests to be certain that I'm teaching her and her classmates what they really need to learn."

Now, how would you respond to Mrs. Smothers?

CHAPTER SUMMARY

In this chapter, the emphasis was on why teachers really need to know about assessment. Early in the chapter, the assessment-related features of various reauthorizations of the Elementary and Secondary Education Act of 1965 were briefly described because this oft-revised federal law's impact on most teachers' instructional and assessment decisions is becoming profound. *Educational assessment* was defined as a formal attempt to determine students' status with respect to educational variables of interest. Much of the chapter was devoted to a consideration of why teachers must become knowledgeable regarding educational assessment. Based on

teachers' classroom activities, four traditional reasons were given for why teachers assess—namely, to (1) determine students' current status, (2) monitor students' progress, (3) assign grades, and (4) determine a teacher's own instructional effectiveness. Based on recent uses of educational assessment results, three more current reasons teachers need to know about instruction were identified. Those more recent functions of educational tests are to (1) influence public perceptions of educational effectiveness, (2) help evaluate teachers, and (3) clarify teachers' instructional intentions. Regardless of the specific application of test results, however, it was emphasized that teachers should use the results of assessments *to make better decisions.* That's really the only excuse for taking up students' time with assessment.

The chapter identified three major outcomes to be attained by those reading the book—namely, becoming more knowledgeable about how to (1) construct and evaluate their own classroom tests, (2) interpret results of standardized tests, and (3) teach students to master what's assessed in classroom and high-stakes tests. It was also suggested that an assessment-literate teacher should attempt to promote parents' and students' assessment literacy.

Determining Your Outcome Mastery

This is your first opportunity to decide how well you have achieved the chief chapter outcome for one of the chapters in *Classroom Assessment.* You will have 15 additional opportunities to do the very same thing as you spin through each of the book's 16 chapters.

Remember, although other topics were addressed in Chapter 1, the most important intended learning outcome was that you would become conversant with what underlies the reasons teachers test their students. As a reminder, here is the chapter's chief 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*

To engage fully in the formative-assessment process, complete both Mastery Checks, reach a mastery-determination judgment and, thereafter, make an adjustment decision. This simple assessment-governed process exemplifies how students employ formative assessment. For teachers, the process is similar, but when teachers review their students' performances on carefully selected tests, those teachers will use students' performances to decide whether to make any instructional adjustments.

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

MyEdLab *Selected-Response Check for Outcome Mastery*
MyEdLab *Constructed-Response Check for 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*

References

Andrade, H. (Ed.). (Winter 2009). "A special issue on classroom assessment," *Theory into Practice, 48,* no. 1.

Black, P., and Wiliam, D. (1998). "Inside the black box: Raising standards through classroom assessment," *Phi Delta Kappan, 80,* no. 2 (October): 139–148.

Curwin, R. L. (2014). "Can assessments motivate?" *Educational Leadership, 72,* no. 1 (September): 38–41.

Goodwin, B. (2014). "Better tests don't guarantee better instruction," *Educational Leadership, 71,* no. 6 (March): 78–81.

Gotch, C. M., and French, B. F. (2014). "A systematic review of assessment literacy measures," *Educational Measurement: Issues and Practice, 33,* no. 2 (Summer): 14–18.

Hess, F. M. (2008/2009). "The new stupid," *Educational Leadership, 66,* no. 4 (December/January): 12–17.

Layton, L. (June 7, 2014, Issue). "How Bill Gates pulled off the swift Common Core revolution," *The Washington Post.*

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards for English Language Arts and Mathematics.* Washington, DC: Author.

Nichols, S. L., and Berliner, D. C. (2008). "Testing the joy out of learning," *Educational Leadership, 65,* no. 6 (March): 14–18.

Popham, W. J. (2009). *Unlearned lessons: Six stumbling blocks to our schools' success.* Cambridge, MA: Harvard Education Press.

Stiggins, R. J., and Chappuis, J. (2012). *An introduction to student-involved assessment FOR learning* (6th ed.). Boston: Pearson.

Wiliam, D. (2013). "Assessment: The bridge between teaching and learning," *Voices from the Middle, 21,* no. 2 (December): 15–20.