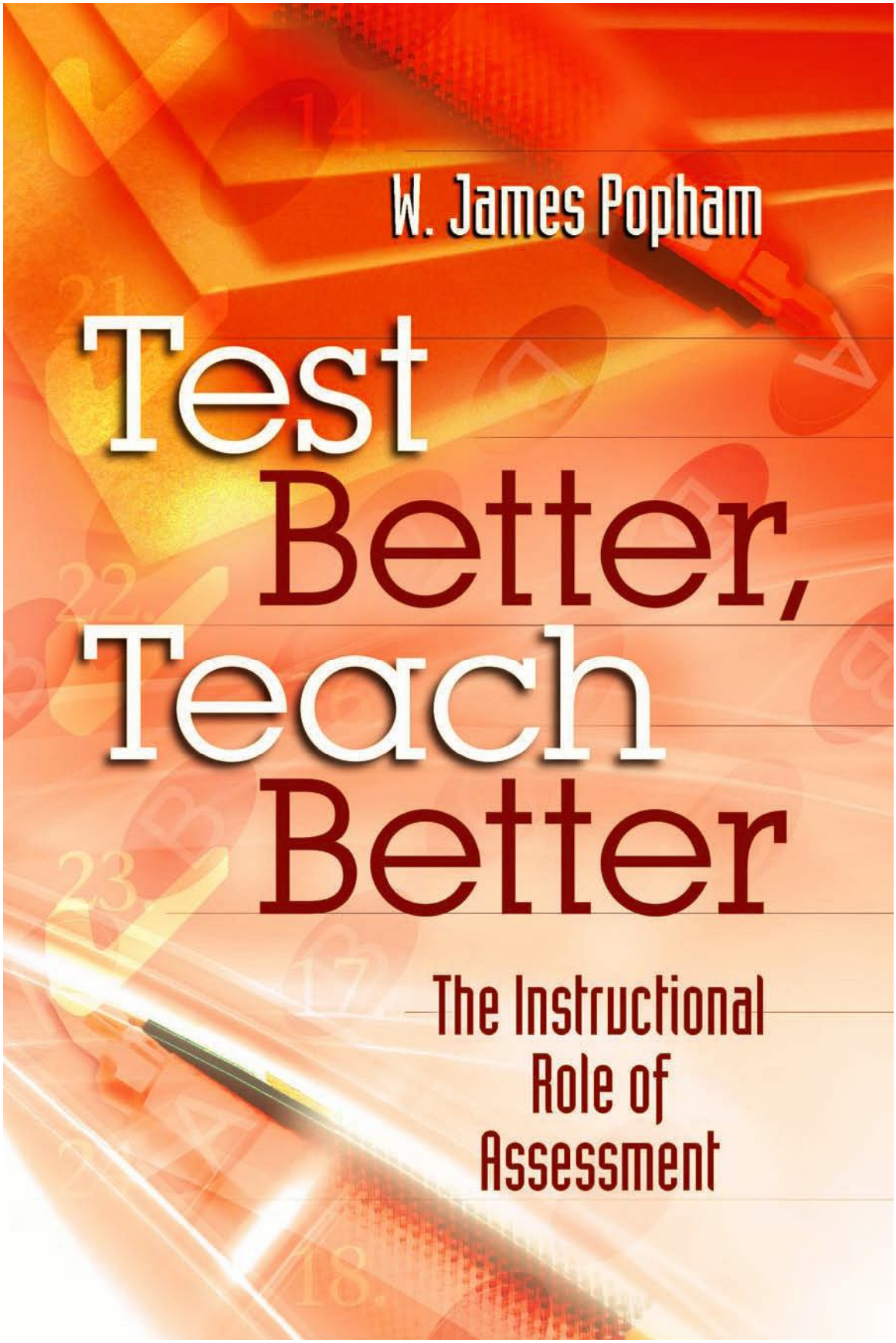

W. James Popham
Test Better, Teach Better
The Instructional Role of Assessment

W. James Popham

Test Better, Teach Better

The Instructional Role of Assessment

Association for Supervision and Curriculum Development • Alexandria, Virginia USA

Association for Supervision and Curriculum Development
1703 N. Beauregard St. • Alexandria, VA 22311-1714 USA
Telephone: 800-933-2723 or 703-578-9600 • Fax: 703-575-5400
Web site: http://www.ascd.org • E-mail: member@ascd.org

Gene R. Carter, *Executive Director;* Nancy Modrak, *Director of Publishing;* Julie Houtz, *Director of Book Editing & Production;* Katie Martin, *Project Manager;* Georgia McDonald, *Senior Graphic Designer;* Valerie Sprague and Keith Demmons, *Desktop Publishers*.

Printed in the United States of America.

Paperback IBSN: 0-87120-667-6 • ASCD product #102088 • List Price: $24.95 ($19.95 ASCD member price, direct from ASCD only) s8/03
e-books ($24.95): netLibrary ISBN 0-87120-882-2 • ebrary ISBN 0-87120-934-9 • **Retail PDF** ISBN 1-4166-0121-X

Library of Congress Cataloging-in-Publication Data (for paperback book)

Popham, W. James.
Test better, teach better : the instructional role of assessment / W. James Popham.
p. cm.
Includes bibliographical references and index.
ISBN 0-87120-667-6 (alk. paper)
1. Educational tests and measurements. 2. Examinations. 3. Effective teaching. I. Title.

LB3051.P61433 2003
371.26'2—dc21
2003011511

12 11 10 09 08 07 06 05 04 12 11 10 9 8 7 6 5 4 3 2

Test Better, Teach Better

The Instructional Role of Assessment

Preface

If you could magically and surreptitiously slip into the back of any American classroom these days, odds are you'd find the students taking a test. At least that's the way it must seem to most of our nation's teachers. More and more frequently, teachers find themselves obliged to give their students state-mandated or district-mandated tests that they work in amongst their own classroom testing practices. "Recess" used to mean a time when students were let out to play. Now, at least to many teachers, it's a blessed week in which teachers are not required to administer some sort of externally imposed test.

Accountability Pressures and a New Federal Law

The chief reason for what seems to be an explosion of educational testing is that U.S. educational policymakers, bent on making the nation's educators more accountable, want hard evidence regarding how well public schools are performing. These policymakers, and most of our citizens as well, believe that student test performance should be the ultimate yardstick by which we measure a school's effectiveness. Naturally, then, teachers are under pressure to raise their students' test scores. You know the logic: High test scores signify good schooling and low test scores signify bad schooling.

The already thriving national obsession with educational testing intensified in early 2002, when President George W. Bush signed the

No Child Left Behind (NCLB) Act, an enormously significant piece of federal legislation laced with loads of assessment-and-accountability provisions. The most widely publicized of these, set to take effect in the 2005–2006 school year, is a requirement for every state to conduct achievement testing in reading and mathematics: once a year for all students in grades 3–8 and at least once again in high school, prior to graduation. In addition, NCLB also requires every state to carry out science assessments in several grade ranges, beginning in the 2007–2008 school year. Given the availability of the data these tests will generate, it's certain that NCLB-sired test results will play a prominent role in state-level accountability systems that associate student scores with school and teacher quality.

From a teacher's perspective, however, it is difficult to figure out why more mandated testing will lead to a higher quality of instruction—especially as current accountability-spawned educational tests results have little utility *in the classroom*. In fact, many teachers find that the instructional benefits provided by today's required educational tests are almost nonexistent. Too rarely do teachers find that a student's performance on these tests helps diagnose that student's strengths and weaknesses. Too rarely do teachers find that a student's test performance gives them a better idea of what it is they need to do instructionally to help that student achieve better results. The result of this high pressure with little educational benefit is that more and more teachers associate *testing* with something negative—something to be dreaded or "dealt with" rather something to be embraced as illuminating, helpful, and even essential to better learning. And yet, if properly conceived, educational testing *is* illuminating, helpful, and even essential to better learning. In the pages that follow, I'll try to show you why.

The Intended Medicine

Too much testing has become a sort of sickness in some schools. But tests—the right kinds of tests—*can* give teachers really powerful

insights about how best to teach their students. And teaching students, of course, is the reason most folks chose to become teachers in the first place. Yet, the distressing reality is that teachers who do not possess at least rudimentary knowledge about *testing* are less likely to do a solid job of *teaching*. And that's what this book is about: the kind of testing that improves one's teaching.

I wrote this book for three types of readers:

- Experienced teachers who were not required to complete coursework in educational testing during their preservice teacher-education days. There are many such teachers.
- Experienced teachers who may have taken a course in educational measurement but found the course's theoretical orientation unrelated to the real-world travails of the classroom. There are lots of these teachers as well.
- Students in teacher-education programs who, perhaps in connection with a course in instructional methods or educational psychology, have been asked by their instructor to take a serious look at educational testing. I *hope* there are many such teacher-education students!

All right, now that I've told you for whom I wrote the book, let me tell you why I wrote it and what you'll find here. I wrote this book to inform educators, and those preparing to be educators, about some basic things they need to know regarding educational testing. More specifically, I want to help teachers master a set of measurement-related skills and knowledge they'll need if they are going to teach their students effectively.

Instructionally Focused Testing: Two Species

There are really two kinds of educational tests that may (or may not) help teachers do a better instructional job. The first is a teacher's *classroom tests,* typically designed by the teacher to measure student

mastery of specific unit content. The second is *externally imposed tests,* those tests required by state or district authorities and designed by professional test developers to measure student mastery of the sets of objectives experts have deemed essential.

Because teachers obviously have far more influence over tests they create for their own students, I've devoted the majority of this book to the innards of teacher-made classroom tests and the measurement concepts that have a direct bearing on classroom test construction and use. But, given the test-obsessed reality most teachers are living in, I have also addressed several issues related to external testing. Teachers must become more familiar with the uses and misuses of externally imposed tests so that they can recognize when an unsound test is being forced on their students, protest persuasively against such testing, and, over time, influence these tests' revision. Thus, as you read through this book, you'll see that some content pertains exclusively to classroom tests, some content pertains only to external tests, and some content pertains to both.

Two Books About Testing: Comparing and Contrasting

Not so long ago, I wrote a book called *The Truth About Testing*, also published by the Association for Supervision and Curriculum Development (ASCD). Because both books deal with educational testing, it's possible that some confusion could arise about the two books' treatment of the topic. I thought a few sentences dealing directly with this matter might be helpful.

The Truth About Testing is subtitled *An Educator's Call to Action.* I wrote it to inform educators of several assessment-induced problems that I believe are eroding the quality of schooling in the United States. I tried to explain those test-related problems, and then I laid out a series of actions educators could take to deal with the situation. In short, *The Truth About Testing* is an experience-based entreaty to my educational colleagues (influential educational leaders especially) to *do something* to fix the problems resulting from the use of the

wrong kinds of tests in our schools. In contrast, the book you're reading now represents an attempt to familiarize teachers with the kinds of assessment practicalities they need so that they can get the most out of their own classroom tests and ensure, insofar as possible, that any externally imposed assessments their students take are educationally defensible.

Do the two books overlap? Yes, at certain points they do, because there are some assessment-related concepts that are so significant that they need to be understood by *every educator*. For example, in both books, I've tried hard to get readers to realize why it is that traditionally constructed standardized achievement tests are having such an adverse affect on U.S. education. I believe that both current and future teachers need to understand the reasons underlying this problem. However, *The Truth About Testing* gave little if any attention to the creation and use of the kinds of classroom tests that I describe in this book. That's because classroom teachers really do need to know how to provide care and feeding for varied sorts of test items. The current book, then, is not a "call to action," but a guide to better test use.

A Plunge into the Pool of Practical Assessment

With the *why-I-wrote-this* covered, let's turn to content. I promise that everything about testing that you will read in this book will have a direct bearing on classroom-based instructional decision making. In fact, to keep me honest, and to help you tie a mental ribbon around each chapter, I've concluded each of the book's 11 chapters with a set of "Instructionally Focused Testing Tips," a few concise points clarifying the most salient implications for classroom teachers.

I'd like to make one last point before you proceed to the rest of the book. This is a short book, a book deliberately written for busy people. Instead of providing a comprehensive treatment of these topics, I've synthesized the important ideas that focus on *what matters most* to teachers faced with instructional decisions. As you read

through my nuts-and-bolts coverage, you are almost certain to encounter some topics about which you'd like to learn more. Obligingly, at the end of each chapter, I've provided a set of recommended resources to guide more detailed excursions into the topics treated briefly here.

You can think of this book as a crash course in instructionally focused testing. If you are an experienced teacher, you'll be able to roll through the entire thing in an evening or two. If you are a student in a teacher-education program, then read what your professor says to read! But regardless of your reading rate, I can assure you that mastery of the concepts I present here will help you teach better. And if you teach better, then your students will learn better.

Faced with such persuasive logic, how can you wait any longer to get started?

1

The Links Between Testing and Teaching

YOU'D PROBABLY FIND IT DIFFICULT TO LOCATE ANYONE, TEACHER OR NON-teacher, who doesn't recognize that there's *some* sort of a relationship between teaching and testing. Just about everyone realizes that if a teacher does a great instructional job, that teacher's students will usually perform better on tests. It's the other side of the equation that's less often understood, namely, that *how* a teacher tests—the way a teacher designs tests and applies test data—can profoundly affect *how well* that teacher teaches.

The connection between one's teaching and one's testing is a critical one that, if properly understood, can lead to a substantial increase in instructional effectiveness. I want you not only to accept the idea that testing can help teaching, but also *to act* on that idea. I want you to pick up tangible instructional payoffs from linking your tests to your teaching. You'll teach better, and your students will learn more. You'll be a better teacher, and I'll be a happy author. Let's get started.

What's in a Name?

I need to define some terms as we get under way. First, what is a *test* or, more specifically, what is an *educational test?* Simply put, an educational test is a formal attempt to determine a student's status with

respect to specific variables, such as the student's knowledge, skills, and attitudes. The adjective "formal" in the previous sentence is important, because it distinguishes a test from the many casual judgments that teachers routinely make about their students. For example, during my first year of teaching (in a small eastern Oregon high school), I had a student named Mary Ruth Green. I could almost always tell (or so I thought) how well Mary Ruth had mastered the previous night's English homework assignment. When it came time to discuss the homework topic, if Mary Ruth was animated and eager to contribute, I concluded that she knew the assigned stuff. If she sat silently and avoided eye contact with me, however, I guessed that she and the previous night's homework topic were unacquainted.

I made all sorts of on-the-spot judgments about what Mary Ruth and my other students knew, but those judgments were informal ones and often based on pretty skimpy observational data. In contrast, a test entails a systematic effort to get a fix on a student's status with respect to such things as the student's ability to perform an intellectual skill—to compose a job-application letter, for instance, or to carry out an hypothesis-testing experiment in a chemistry class.

For many people, the word *test* conjures up images of traditional, paper-and-pencil forms (multiple-choice exams or True-False quizzes). Perhaps this explains why a growing number of educators prefer to use the term *assessment,* which seems to embrace both traditional forms of testing and comparatively recent ones like looking for evidence of learning by examining student-generated work portfolios or group reports of experimental projects. Still, as long as you don't restrict yourself to only traditional testing approaches, the terms *test* and *assessment* are really interchangeable. And while we're swimming in this particular synonym pool, let me toss in two more: the slightly technical-sounding *measurement* and the serious-sounding *examination* (or *exam*). Each of these four terms describes a formal attempt to determine a student's status with respect to an educationally relevant variable. In this book, you'll find that I use all four

interchangeably, not for any subtle reasons, but just because I get tired of using the same word all the time.

Why We Test

Human beings are tough to figure out. Ask any psychiatrist. Ask yourself. And young human beings in grades K–12 are no exception. To illustrate, if a teacher wants to determine what Ted's ability to read is, the teacher won't find that information tattooed on Ted's arm. Ted's reading ability is *covert*. The teacher must figure out how to uncover that hidden ability. So the teacher whips up a 15-item reading test calling for Ted to read several short passages and then answer a series of questions getting at (1) the central messages in the passages and (2) certain key details in those passages. Ted takes the test and does a great job, answering each of the 15 items correctly. The teacher then makes an *inference* about Ted's covert reading ability based on Ted's overt performance on the 15-item test.

If you think about it, just about every worthwhile thing that educators try to promote is unseeable. Consider spelling ability as another example. A child's spelling ability cannot be seen, only inferred. What goes through the teacher's head is something like this:

> Martha did well on this month's spelling test. She wrote out "from scratch" the correct spellings for 18 of 20 words I read out loud. It is reasonable for me to infer, then, that Martha possesses a really high level of spelling ability—a level of ability that would display itself in a fairly similar fashion if Martha were asked to take other, similar 20-item spelling tests.

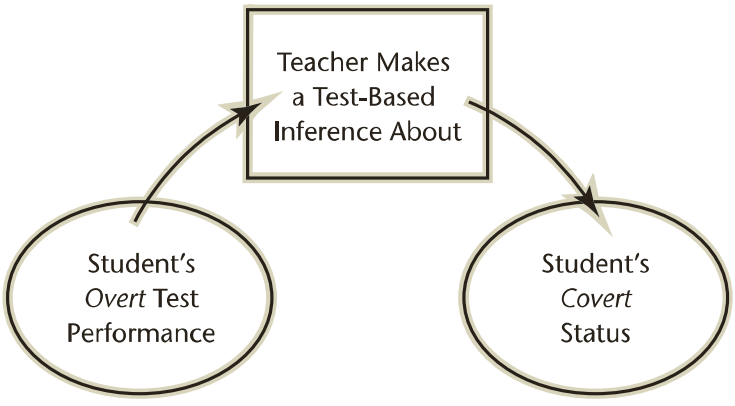
Remember, what the teacher sees when Martha spells the word "awry" properly is only Martha's spelling of "awry" and *not* Martha's spelling ability. The teacher needs to *infer* the level of Martha's spelling skill by seeing how well Martha does on her spelling tests. The more spelling tests that Martha takes, the more confidence the

teacher can have in any inferences about Martha's spelling skill. An inference about a student can be based on a single test; a more accurate inference will be made if multiple tests are employed.

Likewise, a child's ability to perform basic arithmetic skills is unseeable; it's something we infer from the child's performance on an exam (or, preferably, more than one exam) dealing with adding, subtracting, multiplying, and dividing. Children's confidence in being able to present an oral report to their classmates is certainly unseeable, but again, we can infer it from students' responses to an assessment instrument constructed specifically to measure such things. (You'll learn more about that sort of noncognitive assessment in Chapter 8.)

So educational measurement is, at bottom, *an inference-making enterprise* in which we formally collect overt, test-based evidence from students to arrive at what we hope are accurate inferences about students' status with respect to covert, educationally important variables: reading ability, knowledge of history, ability to solve simultaneous equations, interest in social studies, and so on. The process is represented in Figure 1.1.

1.1 **EDUCATIONAL TESTING AS AN INFERENCE-MAKING PROCESS**



Yes, as my experience with Mary Ruth and her homework showed, it is certainly possible for a teacher to make an inference about students based on informal, nontest evidence. Suppose your student Alvin gives you a note in which he had misspelled several words. Based on this evidence, you might infer that Alvin's spelling ability isn't all that wonderful. However, a *formal* assessment of Alvin's spelling skill, one based on a larger and more age-appropriate set of words, would increase the likelihood of your making an accurate inference about Alvin's spelling ability.

The accuracy of these inferences is critical, because a teacher's understanding of students' knowledge, abilities, and attitudes should form the basis for the teacher's instructional decisions. And, of course, the more accurate the test-based inferences a teacher makes, the more defensible will be the teacher's instructional decisions based on those inferences.

What Sorts of Teaching Decisions Can Tests Help?

I've been touting the tight relationship that should be present between testing and teaching. It's time to get more specific. There are four types of teaching decisions that should rest squarely on what a teacher finds out either from the structure of the educational tests themselves or from the way students perform on educational tests.

Decisions about the nature and purpose of the curriculum. Essentially, the teacher seeks answers to questions like these: "What am I really trying to teach? What do my students need to know and be able to do? How can I translate the big curricular goals set for my students into specific, teachable components?"

Decisions about students' prior knowledge. Questions include, "What do my students already know about the topic I'm planning to teach? Are there any gaps that I need to address before we can tackle this material? Based on what my students know and can do, how can I tailor my instruction to provide the proper balance of remediation and challenge?"

Decisions about how long to teach something. Questions include, "How long do I think it will take my students to master this content? What kind of progress are they making? Are we on the right track? Should I continue teaching on my planned schedule, or are we ready to move on?"

Decisions about the effectiveness of instruction. Questions include, "Did my students learn? Was the instructional approach I took a good one? What specific activities were the most advantageous? Where do I need to make alterations?"

Now, let's take a closer look at how tests—both their design and the results of their application—can help teachers make these kinds of decisions with confidence.

Using Tests to Clarify the Curriculum

Typically, educators think of a curriculum as the set of intended outcomes that we want students to achieve. During the bulk of my teaching career, most teachers have used the phrase *educational objectives* to describe their curricular intentions. These days, of course, we find that most curricula are described as sets of *content standards*—that is, the knowledge and skills students are supposed to master as a consequence of instruction. Sometimes we see the term *benchmarks* used to describe the more specific skills and knowledge often subsumed beneath fairly broad content standards. The descriptors may change, but the mission of a curriculum remains constant: Its essential purpose is to lay out the stuff we want kids to learn.

Regardless of whether we call them content standards, goals, or objectives, the curricular intentions handed down by states and districts are often less clear than teachers need them to be for purposes of day-to-day instructional planning. For example, a group of elementary teachers might find themselves responsible for promoting this district-approved social studies content standard: "Students will comprehend the formal and informal nature of the interrelationships

among the executive, legislative, and judicial branches of U.S. government."

Let's imagine you're one of the 5th grade teachers who is supposed to help students master this content standard. How would you go about planning your instruction? Personally, I think there's way too much fuzz on this curricular peach. Different teachers could easily read this social studies content standard and come up with quite divergent ideas of what it signifies. For example, one teacher might conclude that this content standard focuses exclusively on the formal and informal "checks and balances" when one governmental branch interacts with the other two. Another teacher might think that this content standard emphasizes the distinction between "formal" and "informal" interrelationships among the three governmental branches.

Now suppose that your 5th graders will be taking an important "standards-based" social studies achievement test at the end of the school year. If the people who built that test interpret this social studies content standard in one way, and you interpret it in another way—and *teach toward your interpretation*—it's almost certain that your students won't do as well on the achievement test as you, your principal, or your students' parents would like.

Clearly, if the curricular aims that a teacher must address are open to multiple interpretations, then off-the-mark instruction is likely to occur, bringing with it lower test performances. But if a curricular goal is accompanied by a set of illustrative test items indicating the ways that the goal will be measured, then teachers can analyze those items and form a far more accurate idea of the outcome that the state or district is actually seeking. Because the sample test items exemplify what the curricular intention really represents, teachers can plan and provide their students with better, more curricularly relevant instruction.

To illustrate, suppose you knew that mastery of the fairly fuzzy 5th grade social studies goal about the three branches of the U.S. government would be assessed by items similar to the following:

Sample Item 1

Which of the following three branches of U.S. government, if any, is primarily responsible for the final enactment of treaties with foreign nations?

a. Legislative
b. Executive
c. Judicial
d. No single branch is responsible.

Sample Item 2

Which, if any, of the following statements about governmental stability is true? (Mark each statement as True or False.)

a. The enactment of term-limiting legislation at the local level has made the U.S. federal legislative branch of government more stable.
b. The availability of the impeachment process tends to decrease the stability of the executive branch of U.S. government.
c. Historically, the judicial branch of U.S. federal government has been the most stable.

Sample Item 3

Our founding fathers charted a meaningful series of governmental checks and balances. Focus on the area of **taxation,** then select two of the three branches and briefly describe the *formal* way(s) in which one branch can check the other. Answer in the space provided below.

_____

_____

_____

Having read these sample items, wouldn't you have a much better idea of what to teach your students in order for them to come to "*comprehend the formal and informal nature of the interrelationships*

among the executive, legislative, and judicial branches of U.S. government"? Sample Item 1 makes it clear that students will need to learn the primary responsibilities of each governmental branch. Sample Item 2 suggests that students must learn *why* important factors such as governmental stability are present for each branch. And Sample Item 3 indicates that, as the content standard said, students will need to understand the "formal and informal nature of the relationships" among the governmental branches. For this item, as you can see, the focus is on *formal*. In another item, you can reasonably assume, the focus might be on *informal*. Moreover, Sample Item 3 tips you off that students may need to display this understanding by constructing their own responses, rather than merely selecting a response from a set of options.

I believe that elementary teachers who consider these three illustrative items along with the original statement of the content standard are going to have a far more lucid idea of what the content standard actually means. Consequently, they'll be able to deliver instruction that is more on-target and more effective.

The payoffs from test-triggered clarity about curriculum goals can apply with equal force to a teacher's own, personally chosen curricular aspirations. If teachers are pursuing curricular aims of their own choosing, but those aims are less clear (in a teacher's mind) than is desirable for instructional planning purposes, then teachers are likely to come up with less relevant instruction. To illustrate, when I was a first-year teacher, I wanted the students in my two English classes "to be better writers." But even though that very general goal was in my mind as the school year got under way, I really had no idea of what it meant for my students to be "better writers." As the months went by, I occasionally had my students write a practice essay. However, for their final exam, I had them answer multiple-choice items about the mechanics of writing. Shame on me!

The task of creating a few sample assessment items can bring the desired outcomes into focus. In short, test-exemplified curricular

goals will almost always be better promoted instructionally than will unexemplified curricular goals. Because of the significance of tests in helping teachers clarify their instructional targets, I'm going to dig into this topic a bit more deeply in Chapter 2. Stay tuned.

Using Tests to Determine Students' Entry Status

In most instructional settings, teachers inherit a new crop of students each year, and more often than not, these teachers really don't know what sorts of capabilities the new students bring with them. Likewise, teachers looking ahead in their planning books to new topics or skills (weather systems, Homer's epics, multiplying fractions, group discussion skills, ability to work independently) frequently find they have only the roughest idea, usually based on the previous grade level's content standards, of their students' existing familiarity or interest in the upcoming topics or of their students' expertise in the upcoming skill areas. Knowing where students stand in relation to future content, both as a group and as individuals, is one of a teacher's most valuable tools in planning appropriate and engaging instruction. Therefore, it's an eminently sensible thing for teachers to get a fix on their students' entry status by pre-assessing them, usually using teacher-created tests to find out what sorts of skills, knowledge, or attitudes these students have. The more diagnostic a pretest is, the more illuminating it will be to the teacher.

You can use pretests to isolate the things your new students already know as well as the things you will need to teach them. If you are a middle school English teacher aspiring to have your 8th graders write gripping narrative essays, and you're certain that these 8th graders haven't seriously studied narrative essays during their earlier years in school, you could use a pre-assessment to help you determine whether your students possess important *enabling* subskills. Can they, for example, write sentences and paragraphs largely free of mechanical errors in spelling, punctuation, and word usage? If their pre-assessment results show that they already possess these enabling subskills, there's no need

to *re-teach* such subskills. If the pre-assessment results show that your students' mastery of the mechanics of writing is modest, then you'll need to devote appropriate time to promoting such subskills before you move on.

This example brings up an important point. If you're using a classroom pretest chiefly to get a picture of what your students already can do regarding a particular content standard, you should always try to employ a pretest that covers the standard's key enabling subskills or bodies of knowledge. For instance, when I taught a speech class in high school, I always had my students deliver a two- to three-minute extemporaneous speech early in the term. I was looking particularly for the fundamentals—posture, eye contact, organization of content, introductions, conclusions, and avoidance time-fillers such as "uh" and "you know"—those things I knew students needed to master before they could work on refining their abilities as first-class public speakers. Those pretests helped me decide where I wanted to aim my early instruction, and it was always at the most serious weaknesses the students displayed during their "mini-orations."

Using Tests to Determine How Long to Teach Something

One of the classes I taught in my early years on the "grown-up" side of the desk was 10th grade geography. Thanks to a blessed red geography textbook and my ability to read more rapidly than my 10th graders, I survived the experience (barely). I remember that one of my units was three-week focus on map projections and map skills, during which we explored the use of such map-types as Mercator and homolographic projections. Each year that I taught 10th grade geography, my three-week unit on maps was always precisely three weeks in length. I never altered the duration of the unit because, after all, I had originally estimated that it would take 15 days of teaching to stuff the designated content into my students' heads. Yes, I was instructionally naïve. Beginning teachers often are.

What I should have done instead was use some sort of "dipstick" assessment of students' map skills throughout that three-week period to give me a better idea of how long I really needed to keep teaching map skills to my 10th graders. I always gave my students a 30-item map skills exam at the end of the 3 weeks; I could easily have taken that exam and split it up into 15 microquizzes of 1 or 2 items each, and then randomly administered each of those microquizzes to different students at the end of, say, 2 weeks. Students would have needed only two or three minutes to complete their microquizzes.

This approach is a form of what's called *item sampling*, a manner of testing in which different students are asked to complete different subsamples of items from a test. It works quite well if a teacher is trying to get a fix on the status of an entire class. (Clearly, item sampling wouldn't permit sensible inferences about individual students because different students would be completing different microquizzes.) By reviewing the results of my item-sampled, *en route* assessment, I could have determined whether, at the end of only two weeks, my students had already learned enough from their meanderings through Mapland. Looking back, I suspect, we continued to mess with Mercators and homolographics well beyond what was necessary.

You can do something similar with your own students to help you decide how long to continue teaching toward a particular content standard. By using an occasional *en route* test (either item sampling or by giving the same, possibly shortened, test to all of your students), you can tell whether you need to keep banging away on a topic or can put your drumsticks away.

This kind of instructionally illuminating testing, sometimes referred to as *formative assessment,* is a particularly valuable tool today, when there's so much to fit into each school year. The time saved in an easily mastered unit can be time applied to other material that students have unexpected difficulty with. Flexible, *en route* test-guided instructional scheduling can allow your students to move on to fascinating application activities or delve more deeply into other content.

Using Tests to Determine the Effectiveness of Instruction

The issue of how well a teacher has taught is becoming more and more critical as the educational accountability movement places teachers under ever-intensifying scrutiny. The folks who are demanding evidence that teachers are doing a solid instructional job are looking for hard evidence that proves instructional effectiveness.

This is such a serious and complicated assessment-related issue that I've devoted three chapters in this book to it. Chapter 9 explores how *not* to evaluate a teacher's effectiveness; Chapters 10 and 11 tell how to go about it properly. But, because finding out how effective your own instruction is should be important to you, I need to address some important assessment-related topics first.

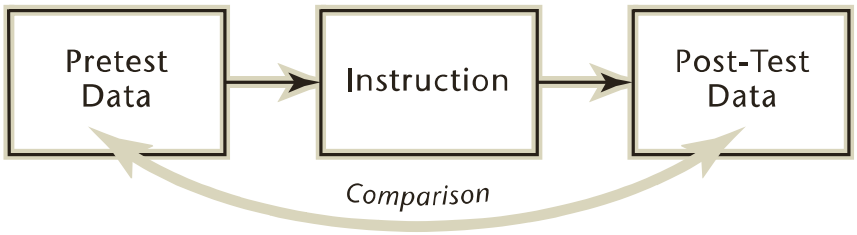
These days, many teachers' instructional competence is being determined on the basis of a single achievement test administered to students each spring. For instance, a 4th grade teacher's students complete a state-approved standardized achievement test in May, and then the test results of this year's 4th graders are compared with the test results of last year's 4th graders. If this year's scores are better than last year's scores, the 4th grade teacher is thought to be doing a good instructional job . . . and vice versa.

But this sort of teacher-appraisal model flunks on several counts. For one thing, it relies on the wrong kind of measurement tool, as you'll learn when you get to Chapter 9. And there's another, more obvious shortcoming in these year-to-year comparison models. The problem is that each year's testing takes place with *a different group of students,* and the results depend on the collection of kids being compared. If your students last year were an atypical collection of gifted girls and boys and this year's crop is a more typical mix, then you can expect your year-to-year results to show a decline, regardless of your abilities as an instructor.

The simple little model of pre-assessment and postassessment comparison displayed in Figure 1.2 is the most fundamental way teachers can judge their own teaching skill. A pretest gets a fix on

students' status before instruction (at the start of school year, say) and a post-test measures the same students' status again, once instruction is complete (at the end of the school year).

1.2 A Common Model for Determining Instrumental Impact



As you can see from the figure, the heart of this evaluative model is students' test performance. Although a *teacher's* overall performance should be determined using a variety of evaluative considerations, not just students' test data, one overridingly important factor should be how well the teacher's students have learned what they were supposed to learn. A pretest/post-test evaluative approach (using some refinements that you'll read about in Chapter 11) can contribute meaningfully to how teachers determine their own instructional impact.

Okay, we've considered four ways in which testing—the tests themselves and the student results they produce—can help a teacher make better instructional decisions. The rest of this book will provide you with sufficient information about these and other ways of using assessment in your own classes to make your instructional decisions more defensible.

Instructionally Focused Testing Tips

- Recognize that students' overt responses to educational tests allow teachers to make inferences about students' covert status.
- Use tests to exemplify—and, thus, clarify—fuzzy statements of curricular aims.
- Pre-assess any new group of students to identify those students' entry status. Also pre-assess students when they'll be encountering new skills and knowledge to be learned.
- Use test results to determine how much instruction on a given topic your students need.
- Include the data generated by educational tests in evaluations of your own instructional effectiveness.

Recommended Resources

Falk, B. (2000). *The heart of the matter: Using standards and assessment to learn*. Westport, CT: Heinemann.

Popham, W. J. (Program Consultant). (1996). *Improving instruction through classroom assessment* [Videotape]. Los Angeles: IOX Assessment Associates.

Popham, W. J. (2001). *The truth about testing: An educator's call to action*: Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (Program Consultant). (2002). *Educational tests: Misunderstood measuring sticks* [Videotape]. Los Angeles: IOX Assessment Associates.

Ramirez, A. (1999, November). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan*, *81*(3), 204–208.

Shepard, L. A. (2000, October). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Sirotnik, K. A., & Kimball, K. (1999, November). Standards for standards-based accountability systems. *Phi Delta Kappan*, *81*(3), 209–214.

Stiggins, R. J. (Program Consultant). (1996). *Creating sound classroom assessments* [Videotape]. Portland, OR: Assessment Training Institute.

Stiggins, R. J. (2001). *Student-involved classroom assessment* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Wiggins, G., Stiggins, R. J., Moses, M., & LeMahieu, P. (Program Consultants). (1991). *Redesigning assessment: Introduction* [Videotape]. Alexandria, VA: Association for Supervision and Curriculum Development.