

W. James Popham

Assessment Literacy for Teachers: Faddish or Fundamental?

In recent years, increasing numbers of professional development programs have dealt with assessment literacy for teachers and/or administrators. Is assessment literacy merely a fashionable focus for today's professional developers or, in contrast, should it be regarded as a significant area of professional development interest for many years to come? After dividing educators' measurement-related concerns into either classroom assessments or accountability assessments, it is argued that educators' inadequate knowledge in either of these arenas can cripple the quality of education. Assessment literacy is seen, therefore, as a sine qua non for today's competent educator. As such, assessment literacy must be a pivotal content area for current and future staff development endeavors. Thirteen must-understand topics are set forth for consideration

by those who design and deliver assessment literacy programs. Until preservice teacher education programs begin producing assessment literate teachers, professional developers must continue to rectify this omission in educators' professional capabilities.

FOR THE PAST SEVERAL YEARS, *assessment literacy* has been increasingly touted as a fitting focus for teachers' professional development programs. The sort of assessment literacy that is typically recommended refers to a teacher's familiarity with those measurement basics related directly to what goes on in classrooms. Given today's ubiquitous, externally imposed scrutiny of schools, we can readily understand why assessment literacy might be regarded as a likely target for teachers' professional development. Yet, is assessment literacy a legitimate focus for teachers' professional development programs or, instead, is it a fashionable but soon forgettable fad?

W. James Popham is a Professor Emeritus of education at the University of California, Los Angeles.

Correspondence should be addressed to W. James Popham, 28170 S.W. Boberg Road, No. 1, Wilsonville, OR 97070. E-mail: wpopham@ucla.edu

The Consequences of Omission

Many of today's teachers know little about educational assessment. For some teachers, test is a four-letter word, both literally and figuratively. The gaping gap in teachers' assessment-related knowledge is all too understandable. The most obvious explanation is, in this instance, the correct explanation. Regrettably, when most of today's teachers completed their teacher-education programs, there was no requirement that they learn anything about educational assessment. For these teachers, their only exposure to the concepts and practices of educational assessment might have been a few sessions in their educational psychology classes or, perhaps, a unit in a methods class (La Marca, 2006; Stiggins, 2006).

Thus, many teachers in previous years usually arrived at their first teaching assignment quite bereft of any fundamental understanding of educational measurement. Happily, in recent years we have seen the emergence of increased pre-service requirements that offer teacher education candidates greater insights regarding educational assessment. Accordingly, in a decade or two, the assessment literacy of the nation's teaching force is bound to be substantially stronger. But for now, it must be professional development—completed subsequent to teacher education—that will supply the nation's teachers with the assessment-related skills and knowledge they need.

Two Types of Assessments

The term *assessment* in this instance should be regarded not as merely a traditional paper-and-pencil test or, for that matter, any kind of formal test. On the contrary, an assessment might consist of a wide variety of evidence-eliciting techniques such as asking students to respond to teacher-presented questions by using individual, erasable white boards during a class discussion, or conducting oral interviews with solo students or with groups of students. Similarly, an assessment of students' attitudes might feature the use of anonymous, self-report inventories. To assess students' cognitive skills, teachers might employ

an elaborate performance test in which students must complete, then describe in writing, a series of independent, scientific experiments. *Assessment*, therefore, should most definitely not be regarded as synonymous with *test*. (See Cizek in this issue for further explanation of the difference between assessment and test.)

If we could magically track a given teacher's total career in the classroom, we would surely encounter innumerable instances in which educational assessment impinged directly on the decisions this teacher needed to make. Some of those assessment-related decisions would be quite important, such as whether a student should be promoted to the next-higher grade. Some of those assessment decisions would be less important, such as whether the teacher should assign high or low marks to students on daily quizzes of modest significance. It is fairly obvious that the seriousness of a teacher's assessment-related decisions can range substantially. But, granting that the significance of teachers' assessment-related decisions can bounce all over the place, a key question remains: *What kinds of assessments do teachers most need to understand?*

Several sensible ways to subdivide our current educational assessment cake are currently available. For instance, one increasingly popular distinction is to contrast *summative assessment* and *formative assessment* (McMillan, 2007). Summative assessment refers to the use of assessment-based evidence when arriving at decisions about already-completed instructional events such as the quality of a year's worth of schooling or the effectiveness of a semester-long algebra course. Summative assessment is intended to help us arrive at go/no-go decisions based on the success of a final-version instructional program.

In contrast, formative assessment is a process in which assessment-elicited evidence is used by teachers to adjust their ongoing instructional activities, or by students to adjust the ways they are trying to learn something. In contrast to its summative sibling, formative assessment has a powerful *improvement* orientation, because it is intended to stimulate ameliorative adjustments in teachers' still-malleable instructional programs or in students' current learning-tactics. For in-

stance, formative assessment would be seen when a teacher frequently administers brief dip-stick quizzes, not for grading purposes but, instead, to let the teacher and the students see whether they need to make any changes in what they are doing in class. The function of the formative assessment process is to supply evidence that will enhance students' learning (Popham, 2008).

Chiefly because of formative assessment's relatively recent arrival on the measurement scene, and because there are still some serious definitional disputes regarding what the precise nature of formative assessment is and is not (McMillan, 2007), for purposes of the following analysis the two kinds of assessments I consider as a framework for the promotion of assessment literacy are *classroom assessments* and *accountability assessments*. *Classroom assessments* refer to those formal and informal procedures that teachers employ in an effort to make accurate inferences about what their students know and can do. Sometimes the results of classroom assessments are employed by teachers to improve an underway instructional program, in much the same way teachers would proceed when classroom formative assessment is in full flower, so that both the teacher and students are routinely relying on assessment-elicited evidence to monitor and, if necessary, adjust their activities.

More often than not, however, the results of classroom assessments are used simply to dole out grades to students. Although most classroom assessments are constructed by teachers themselves, in some settings we see teachers employing tests they find in a textbook or in the instructor's manual accompanying that textbook. Then, too, some districts—or even some state departments of education—supply teachers with testing instruments or assessment procedures that teachers may, if they wish, employ as classroom assessments.

Accountability assessments are those measurement devices, almost always standardized, used by governmental entities such as states, provinces, or school districts to ascertain the effectiveness of educational endeavors. In the United States, the most prevalent accountability assessments are currently those required by the

No Child Left Behind Act (NCLB), a powerful federal law calling for the administration of reading and mathematics accountability tests in grades 3–8 and once during high school. Although officials of each state possess a certain degree of autonomy in choosing the tests to be used in satisfaction of NCLB requirements, the federal government still exerts considerable pressure on states to employ relatively traditional standardized testing instruments when satisfying the various NCLB regulations and guidelines dispensed by the U.S. Department of Education. Based on students' performances on these accountability tests, a number of far-reaching decisions regarding schools and school districts are reached. To illustrate, if a particular school's staff is unable to improve its students' scores on the state's accountability tests for several years in a row, then the school might be substantially restructured, or even permanently closed down.

Two Decision Clusters

These days, there are two collections of assessment-related decisions that teachers need to make. Unsurprisingly, those decisions are linked to the two major categories of assessments previously identified, namely classroom assessments and accountability assessments. One of those clusters of decisions is fairly obvious, that is, those decisions related to teachers' use of classroom assessments such as which ones to employ and how they should be used. But another, more subtle set of decisions stems from the nature of accountability tests that, almost without exception, have a direct or indirect influence on what teachers do in their classrooms. Both of these clusters of decisions have a meaningful impact on the way students end up being taught. Because assessment-literate teachers will typically make better decisions, and because we want students to be better taught, it should be obvious that today's teachers must acquire more assessment literacy—and the sooner, the better.

Let's look, first, at classroom assessments. If teachers are assessment literate, odds are that their classroom assessments will be better, be-

cause those teachers will know not only what it is that constitutes a defensible versus an indefensible assessment, but also what represents an accurate versus an inaccurate interpretation of assessment-elicited data. Classroom assessments, even if teachers use them in a fairly perfunctory fashion, will have at least some impact on the way students are educated. The more importance that the teacher ascribes to classroom assessments, the more profound will be the impact of such assessments on a classroom's day-to-day instructional activities.

For example, suppose a teacher employs frequent classroom assessments as an integral part of a full-blown formative assessment strategy; that is, a regular process wherein the dominant mission of educational measurement is to help teachers spruce up their instructional activities and to help students sharpen up the ways they are trying to learn things. In that situation, it is clear that classroom assessments will play a prominent role in what goes on instructionally. At the other extreme, even in a classroom where tests are used by a teacher only to assign grades to students, flawed tests can soon discourage poorly assessed students who, in turn, may find their academic motivation diminishing because of the teacher's tawdry tests. Less motivated students will usually end up being less well educated.

Because their impact on the education process can range from modest to major, classroom assessments ought to be as good as they can be. Teachers who are genuinely assessment literate will not only know how to create more suitable assessments, but will also be conversant with a wide array of potential assessment options. All teachers assess their students, some more vigorously than others, and almost all of those assessments have an impact on instruction, some more substantial than others. Accordingly, all teachers need to possess sufficient assessment literacy so their classroom assessments are at least satisfactory or, preferably, substantially better.

Turning now to the decisions teachers face regarding accountability assessments, why is it that teachers need assessment literacy to make decisions related to the use of such tests? As indicated above, although it is unarguable that

teachers need to be assessment literate insofar as such literacy bears on decisions related to their classroom assessments, it is not immediately apparent why teachers' assessment literacy is germane to the sorts of tests traditionally used for accountability testing. After all, these are large-scale tests whose nature and nurture are controlled by high-level governmental officials, not classroom teachers. But, unfortunately, therein rests a fallacy that has flourished for far too long.

A major reason for the reluctance of teachers to scrutinize educational accountability tests is that teachers do not think they know enough to evaluate the quality of those tests. In many instances, that is accurate. Few of us like to judge something we know little about. It is not surprising, therefore, that so few of today's high-stakes tests are seriously scrutinized by the very educators who are, themselves, evaluated by students' scores on those tests.

To illustrate the gravity of the situation currently existing regarding accountability testing, in the United States we find that the vast majority of standardized achievement tests being used as NCLB accountability assessments are *instructionally insensitive*; that is, those tests are unable to distinguish between students who have been skillfully taught and those students who have been shabbily taught (Popham, 2007). An instructionally insensitive test, therefore, fails to provide an accurate picture of how well students in a given school have been taught. Typically, such tests merely measure the affluence-level of a school's students. So, if a teacher works in a school serving economically disadvantaged students, even if that teacher is doing a terrific instructional job, odds are that the test performances of the teacher's students on an instructionally insensitive accountability test will not reflect how well the teacher has taught. Large numbers of inaccurately evaluated schools, of course, negate the very raison d'être for accountability laws such as NCLB. But teachers with insufficient assessment literacy will not understand this, because such teachers typically cannot tell the difference between an accountability test that is instructionally sensitive and one that is instructionally insensitive.

Thus, one overriding decision for a teacher in relation to an accountability test is whether to try to improve an inappropriate accountability test or, perhaps, try to replace it altogether. Teachers, separately, in small groups, or as part of a larger coalition, can decide to take aggressive action to educate, then influence, those who install unsound accountability tests. If teachers can exercise the sort of leadership that one should expect from enlightened professionals, then there is a chance that, in time, inadequate accountability assessments can be replaced by better ones. When our accountability assessments are instructionally sensitive, then our accountability programs will have a chance to work the way they are supposed to work.

I have suggested that when we carve up the world of educational testing into two lumps, classroom assessments and accountability assessments, teachers will be faced with clusters of decisions they will need to make with regard to each kind of assessment. Both sets of decisions are important. Both sets of decisions have an impact on how students are educated. But in order to make those decisions more defensibly, teachers need to possess a solid dose of assessment literacy. It seems clear that such literacy should be conveyed to teachers via professional-development activities.

A Quick Content Dip

Professional development programs focused on assessment literacy need to be tailored. Such a program designed for school *administrators* is likely to be similar to an assessment-literacy program for teachers, in the sense that many of the topics to be treated would be essentially identical, but some salient content differences would—and should—exist. To conclude this analysis, I would like to lay out the content that should be addressed—in a real-world, practical manner rather than an esoteric, theoretical fashion—during an assessment-literacy professional development program for teachers. This will only be a brief listing of potential content, but those who are interested in a closer look at possible

content for such programs will find more detailed treatments of potential emphases in the list of references.

Those considering what to include in an assessment literacy professional development program for teachers should seriously consider focusing on a set of target skills and knowledge dealing with the following content:

1. *The fundamental function of educational assessment, namely, the collection of evidence from which inferences can be made about students' skills, knowledge, and affect.* A common misconception among educators is to reify test scores, as though such scores are the true target of an educator's concern. In reality, the only reason we test our students is in order collect *evidence* regarding what we cannot see—understanding, skill development, and so on. Almost all of our educational goals are aimed at *unseeable* skills and knowledge. We cannot tell how much history a student knows just by looking at that student. Thus, we must rely on students' *overt* test performances to produce evidence so we can arrive at defensible inferences about students' *covert* skills and knowledge.
2. *Reliability of educational assessments, especially the three forms in which consistency evidence is reported for groups of test-takers (stability, alternate-form, and internal consistency) and how to gauge consistency of assessment for individual test-takers.* Many educators place absolutely unwarranted confidence in the accuracy of educational tests, especially those high-stakes tests created by well-established testing companies. When educators grasp the nature of measurement error, and realize the myriad factors that can trigger inconsistency in a student's test performances, those educators will regard with proper caution the imprecision of the results obtained on even some of our most time-honored assessment instruments.
3. *The prominent role three types of validity evidence should play in the building of arguments to support the accuracy of test-based interpretations about*

students, namely, content-related, criterion-related, and construct-related evidence. Anytime an educator utters the phrase *a valid test*, that educator is—at least technically—in error. It is not a test that is valid or invalid. Rather, it is the *inference* we base on a test-taker's score whose validity is at issue. Moreover, the types of validity evidence we collect are fundamentally different. As a consequence, for example, classroom teachers need to know that the chief kind of validity evidence they need to attend to should be content-related.

4. *How to identify and eliminate assessment bias that offends or unfairly penalizes test-takers because of personal characteristics such as race, gender, or socioeconomic status.* During the past two decades, the measurement community has devised both judgmental and empirical ways of dramatically reducing the amount of assessment bias in our large-scale educational tests. Classroom teachers need to know how to identify and eliminate bias in their own teacher-made tests.
5. *Construction and improvement of selected-response and constructed-response test items.* Through the years, measurement specialists have been assembling a collection of guidelines regarding how to create wonderful, rather than wretched, test items. Moreover, once a set of test items has been constructed, there are easily used procedures available for making those items even better. Educators who generate tests need to be conversant with the creation and honing of test items.
6. *Scoring of students' responses to constructed-response tests items, especially the distinctive contribution made by well-formed rubrics.* Although constructed-response test items such as essay and short-answer items often provide particularly illuminating evidence about students' skills and knowledge, the scoring of students' responses to such items often goes haywire because of loose judgmental procedures. Teachers need to know how to create and use rubrics, that is, scoring guides, so students' performances on constructed-response items can be accurately appraised.
7. *Development and scoring of performance assessments, portfolio assessments, exhibitions, peer assessments, and self-assessments.* Gone are the days when teachers only had to know how to score tests by distinguishing between a circled *T* or *F* for students' answers to true-false items. Given the current use of assessment procedures calling for students to respond in dramatically diverse ways, today's teachers need to learn how to generate and perhaps score a considerable variety of assessment strategies.
8. *Designing and implementing formative assessment procedures consonant with both research evidence and experience-based insights regarding such procedures' likely success.* Formative assessment is a process, not a particular type of test. Because there is now substantial evidence at hand that properly employed formative assessment can meaningfully boost students' achievement (Black & Wiliam, 1998a), today's educators need to understand the innards of this potent classroom process.
9. *How to collect and interpret evidence of students' attitudes, interests, and values.* When considering the importance of students' acquisition of cognitive versus affective outcomes, it could be argued that inattention to students' attitudes, interests, and values can have a lasting, negative impact on those students. Teachers, therefore, should at least learn how to assess their students' affect so that, if those teachers choose to do so, they can get an accurate fix on their students' affective dispositions.
10. *Interpreting students' performances on large-scale, standardized achievement and aptitude assessments.* Because students' performances are of interest to both teachers and students' parents, teachers must understand the most widely used techniques for reporting students' scores on today's oft-administered standardized examinations, in-

cluding, for example, what is meant by a *scale score*.

11. *Assessing English Language Learners and students with disabilities*. Although most of the measurement concepts that educators need to understand will apply across the board to all types of students, there are special assessment issues associated with students whose first language is not English and for students with disabilities. Because today's educators have been adjured to attend to such students with more care than was seen in the past, it is important for all teachers to become conversant with the assessment procedures most suitable for these subgroups of students.
12. *How to appropriately (and not inappropriately) prepare students for high-stakes tests*. Given the pressures on educators to have their students shine on state and, sometimes, district accountability tests, there have been reports of test-preparation practices that are patently inappropriate. In many instances, such unsound practices arise simply because teachers had not devoted attention to the question of how students should and should not be readied for important tests. They should be prepared to do so.
13. *How to determine the appropriateness of an accountability test for use in evaluating the quality of instruction*. It is not safe to assume that, because an accountability test has been officially adopted in a state, this test is suitable for evaluating schools. More than ever before, educators need to understand what makes a test suitable for appraising the quality of instruction.

All but a few of these 13 content recommendations are applicable to both classroom assessments and accountability assessments. The recommendations regarding the determination of an accountability test's evaluative appropriateness and interpreting students' performances on large-scale, standardized tests, of course, refer only to accountability assessments. Conversely, the recommendation regarding learning about formative assessment procedures clearly deals

with classroom assessments rather than accountability assessments. Beyond those dissimilarities, however, a professional development program aimed at the promotion of teachers' assessment literacy should show how the bulk of the content recommended here has clear relevance to both classroom assessments and accountability assessments.

Of particular merit these days is the use of *professional learning communities* as an adjunct to, or in place of, more traditional professional development activities. Such communities consist of small groups of teachers and/or administrators who meet periodically over an extended period of time, for instance, one or more school years, to focus on topics such as those identified above. If such a group consists exclusively of teachers, then it is typically referred to as a *teacher learning community*. If administrators are involved, then the label *professional learning community* is usually affixed. Given access to at least some written or electronic materials as a backdrop (e.g., Popham, 2006, which is available *gratis* to such learning communities), collections of educators with similar interest can prove to be remarkably effective in helping educators acquire significant new insights.

Fad-Free Focus?

The presenting question that initiated this analysis was whether professional development programs aimed at enhancing teachers' assessment literacy were warranted, either in the short-term or long-term. I identified two sets of teachers' assessment-related decisions that could be illuminated by such programs, namely, those decisions related to classroom assessments and those decisions related to accountability assessments. Although, at the current time, teachers are surely faced with assessment-dependent choices stemming from both of these sorts of assessments, will both types of assessments be with us over the long haul?

The answer to that question is, in my view, an emphatic *Yes*. With regard to classroom assessments, the influential work of Black and

Wiliam (1998a, 1998b) lends powerful empirical support attesting to the learning dividends of instructionally oriented classroom assessment. When classroom assessments are conceived as assessments for learning, rather than assessments of learning, students will learn better what their teacher wants them to learn. Not only is the evidence supporting such a formative approach to classroom assessment demonstrably effective, but there are—happily—diverse ways to implement an instructionally oriented approach to classroom assessment. As the two British researchers point out:

The range of conditions and contexts under which studies have shown that gains can be achieved must indicate that the principles that underlie achievement of substantial improvements in learning are robust. Significant gains can be achieved by many different routes, and initiatives here are not likely to fail through neglect of delicate and subtle features. (Black & Wiliam, 1998a, pp. 61–62)

It appears, then, that teachers who want to be optimally effective ought to be learning about the essentials of classroom assessment for a long while to come.

Turning to accountability assessment, there seems little reason to believe that the demand for test-based evidence of teachers' effectiveness will evaporate—ever. Accountability pressure on educators springs from taxpayers' doubts that their public schools are as effective as they ought to be. It will take decades of consistent educational success stories before the public is disabused of its skeptical regard for public schools. Even if the public were ever to relax its demands for educational accountability evidence, thoughtful educators still ought to insist on the collection of such evidence. That is the kind of requirement that any self-respecting profession ought to impose on itself.

Thus, it seems that assessment literacy is a commodity needed by teachers for their own long-term well-being, and for the educational well-being of their students. For the foreseeable future, teachers are likely to exist in an environment where test-elicited evidence plays a prominent instructional and evaluative role. In such environments, those who control the tests tend to control the entire enterprise. Until preservice teacher educators routinely provide meaningful assessment literacy for prospective teachers, the architects of professional development programs will need to offer assessment-literacy programs. We can only hope they do it well.

References

- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–73.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- La Marca, P. (2006, June). *Assessment literacy: Building capacity for improving student learning*. Paper presented at the National Conference on Large-Scale Assessment, Council of Chief State School Officers, San Francisco, CA.
- McMillan, J. H. (Ed.). (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College Press.
- Popham, W. J. (2006). *Mastering assessment: A self-service system for educators*. New York: Routledge.
- Popham, W. J. (2007). Instructional sensitivity: Educational accountability's dire deficit. *Phi Delta Kappan*, 89(2), 149–155.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Stiggins, R. J. (2006). Assessment for Learning: A Key to Student Motivation and Learning. *Phi Delta Kappa Edge*, 2(2), 1–19.

Copyright of Theory Into Practice is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.