

Chapter 11

BEST PRACTICES IN LITERACY ASSESSMENT

Peter Winograd

Leila Flores-Dueñas

Harriette Arrington

This chapter will:

- Examine the political and conceptual thinking that influences current assessment practices.
- Discuss the guidelines that teachers should follow in the ethical and constructive use of assessments for all of the students in their classrooms.
- Describe some important assessment strategies, including contextual and cultural surveys, student observation, portfolios, and conferencing.
- Analyze some of the trends that will affect the assessment of literacy in the future.

THEORY AND RESEARCH BASE

All, regardless of race or class or economic status, are entitled to a fair chance and the tools for developing their individual powers of mind and spirit to the utmost.

—*A Nation at Risk*, 1983

These reforms express my deep belief in our public schools and their mission to build the mind and character of every child, from every background, in every part of America.

—PRESIDENT GEORGE W. BUSH, January 2001

We begin our chapter with these two quotes because they provide much of the political and practical context for many of the assessment efforts currently under way in schools throughout the United States. We believe that understanding the politics of assessment is as important as understanding the theoretical and research base of assessments themselves, because teachers, particularly teachers who teach literacy, often find themselves caught by conflicting demands when it comes to the best practices in literacy assessment.

What are the best practices in literacy? One way to answer this question is by placing literacy assessment within the larger context of the national debate on the need to reform the entire system of public education in the United States.

In 1983, *A Nation at Risk* was published. Almost two decades later, the No Child Left Behind Act was passed. As our two quotes indicate, these reform efforts focused on two related crises in American education. The

first is the low levels of academic achievement of American students in general, and the need for higher, “world-class” standards of learning, so that the United States can remain economically competitive in the future. The second crisis is the continued low academic achievement of many black, Hispanic, Native American, inner-city, and poor rural students, and the importance of helping *all* children obtain an equitable and effective education.

Over the last 7 years, national attention has also focused on the crisis in the teaching profession. The National Commission on Teaching and America’s Future (NCTAF; 1996), for example, examined the uneven and often dismal quality of the teaching profession, and argued eloquently for the need to improve teacher preparation, teacher recruitment, teacher induction, professional development, and the ways that schools are organized. The NCTAF also argued that, next to the family, what teachers know and are able to do is the most important influence on what students learn, and that recruiting, preparing, and retaining good teachers is the central strategy for improving the nation’s schools. The seriousness of this issue becomes apparent when one considers that approximately 2 million teachers will need to be hired in the next 7 to 8 years.

In an attempt to address these crises in education, policy makers, educators, and other stakeholders have engaged in a variety of reforms since the mid-1980s. In the last few years, we have witnessed a more comprehensive and systemic approach to education reform. Systemic reform, in this context, refers to efforts that include several key components: (1) the promotion of meaningful goals and standards for all students; (2) the increased focus on accountability and assessment of students, teachers, and schools; (3) the alignment of policy approaches and the coordination of a wide variety of educational, economic, social, and health institutions, and stakeholders to support student achievement; (4) the restructuring and aligning of public education governance to support student achievement; (5) the increase in the funding and other resources needed to support student achievement; and (6) the increased focus on the role of the classroom teacher.

The most recent effort to improve public education has been the passage of Elementary and Secondary Education Act, 2001. This act, known as “No Child Left Behind,” is probably the most significant federal legislation dealing with education in the last 20 years. The Act contains a number of sweeping provisions, including many with a particular focus on assessment, teacher quality, and reading. The Act has both its advocates and detractors (who refer to the legislation as “The No Child Left Untested Act”), but it remains to be seen how this federal legislation will actually

impact the practice of teachers in schools. One thing is for certain, however. The political aspects of literacy assessment are likely to increase.

Let us consider, for a moment, some of the latest evidence behind the concerns over excellence, equity, and the importance of teachers. These data come from a number of crucial sources, including the 2000 and 2001 Quality Counts Reports published by *Education Week* (www.edweek.org), *Kids Count Data Book* (2002; www.aecf.org), the Education Trust (www.edtrust.org), the U.S. Department of Education (2002; www.ed.gov), and the National Assessment of Educational Progress (2000; www.nces.ed.gov/naep).

The following are some of the national statistics related to reading and writing:

- Thirty-two percent of fourth-grade students performed at or above the proficient level in 2000. The proficient level is the level identified by the National Assessment Governing Board (NAGB) that all students should reach.
- The reading performance of the nation's fourth graders has remained relatively stable across assessment years. In 2000, the national average scale score of 217 was similar to that in 1992.
- Although the reading performance of the nation's fourth graders has become relatively stable, significant changes are evident at the upper and lower ends of the performance distribution. Higher performing students' scores in 2000 were significantly higher than in 1992. In contrast, lower performing students' scores in 2000 was significantly lower than in 1992.
- In 2000, students who were eligible for the free/reduced-price lunch program had a lower average score than students who were ineligible for the program; 14% of eligible students performed at or above the *proficient* level in comparison to 41% of noneligible students.
- An overall pattern of declining performance is evident in the average writing scores of 11th-grade students from 1984 to 1996.
- The writing performance of eighth-grade and fourth-grade students did not change significantly from 1984 to 1996.

But some other statistics, which are equally important, let us take a look at the larger picture:

- Seventeen percent of children were in poverty in 2000.
- Eleven percent of teens (1.6 million teenagers ages 16–19) were high school dropouts in 2000.
- Nationally, school districts with the highest child poverty rates have \$1,139 fewer state and local dollars to spend per student compared

with the lowest-poverty districts. That translates into a total \$28,475 for a typical classroom of 25 students.

- Nationally, school districts with the highest minority enrollments have \$979 fewer state and local dollars to spend per student compared with the lowest-minority districts. That translates into a total \$24,475 for a typical classroom of 25 students.
- Some 43% of minority children attend urban schools. Most of them attend schools in which half of the students are poor and predominately, often completely, minority.
- In about half of the states with large cities, a majority of urban students fail to meet even minimum standards on national tests.
- The poorest students are at greatest risk. In urban schools in which most of the students are poor, two-thirds or more of children fail to reach even the “basic” level on national tests.
- Twenty-two percent of the classes in high-poverty schools were taught by teachers without a major or minor in the relevant field in 1993–1994.
- Twelve percent of all newly hired teachers enter the workforce without any training at all, and another 15% enter without having fully met state standards.
- Twenty-three percent of all secondary school teachers do not have a college minor in their main teaching field.

These statistics, and those from other reports produced by national, state, or local organizations, are powerful tools in helping us understand what education is about, and what our priorities should be. We see the debates about meaningful goals and standards, the use of multiple indices to measure progress, the disagreements about what these measures really mean, the arguments over resources, and the struggles to reform teaching and learning as a sign of health in American education. Although the process can be frustrating, it does show a country coming to grips with the reality of public education. One way to measure best practices in assessment at the national level, then, is to see how well they motivate and guide us in improving the lives of all our students.

We can also argue that the best practices in literacy assessment are those that help us understand the larger issues, frame important goals, gather multiple kinds of evidence, and engage in rich discussions about how to help children become better readers, writers, listeners, and speakers. To put it another way, no particular assessment is a best practice in and of itself; rather, the quality of assessments lies largely in how wisely they are used. Here are some thoughts about using assessment wisely.

The most effective practices in literacy assessment are those that occur in the classroom between a competent teacher and a confident

student. The most effective practices in literacy assessment occur when teachers and students can work side by side in a trusting relationship that focuses on growth, nurturance, and self-evaluation. The problem we face, however, is nurturing these kinds of trusting relationships in a educational world dominated by the pressures to raise student achievement, as measured by high-stakes accountability systems.

Let us consider these challenges in more detail. One of the most difficult challenges is ensuring that there is an adequate supply of teachers who are well-prepared in the instruction and assessment of literacy, and that these teachers work in contexts that enable them to focus on meeting the needs of their children. This is no small problem when we consider the nation's need for so many new teachers and that, in too many states, prospective teachers may take only one course in reading. What is even more frightening is that some teachers who enter the classroom through alternative routes may take no classes in reading. Professional organizations such as The International Reading Association, the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association have developed thoughtful standards about what teachers should know and be able to do in terms of evaluation and assessment, but it will take serious revision of teacher preparation programs and professional development opportunities to ensure that the majority of teachers can meet these standards.

Moreover, we need to ensure that when new teachers obtain their licenses, they begin work in schools that are better organized for and more supportive of student and teacher success. Asking teachers who have heavy workloads and too little time to engage in more intensive evaluation and instruction with individual children is unrealistic and unfair. We want to stress that we are not blaming teachers. We are fortunate that the majority of teachers are competent, caring, and committed. What we are saying is that if we truly believe that teachers should engage in the assessment of literacy, then we must strengthen the ways teachers are prepared and improve the ways schools are organized.

Another challenge we face is to think more wisely about how to meet the needs of the children in the classrooms. For example, we must understand that when teachers teach reading or writing, they often face issues that are not just educational in nature. The fact that 68% of America's fourth-graders scored below the proficient level on the 2000 National Assessment of Educational Progress (NAEP) reading exam is a social, political, economic, *and* educational problem. Keeping this context in mind is vital, because the best practices in literacy assessment are those that help us understand these broader issues and how to deal with them systemically, so that all students receive an excellent and equitable education. Assessments that focus on a narrow or isolated aspect of literacy

are not as effective as those that give us a richer picture of the whole child and the world in which he or she lives.

We must also understand that assessments are a very limited tool in dealing effectively and fairly with children and the world they live in. Madaus (1994, p. 79) makes this point most eloquently when he cautions us about becoming too enchanted with assessments and overlooking the needs to attend to student health, nutrition, and living conditions; teacher training; and other critical components in the system of education: "There is a danger that technological solutions, such as alternative educational assessment, will blind policymakers and the public to the reality that we Americans cannot test, examine, or assess our way out of our educational problems (National Commission on Testing and Public Policy, 1990)."

The National Council on Educational Standards and Testing (NCEST; 1992, p. 6) raises a similar point when it states, "Particularly for children, who have historically experienced less success in schools, such as the poor, ethnic minorities and students with disabilities, schools should ensure the opportunity to learn as a critical condition for valid and fair use of assessment results." It is important to note that this is not just concern voiced by a few advocates in the field; rather, it is a specific standard for the fair use of educational and psychological tests developed by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1985; AERA, 1999). The issue of fairness in assessment, especially for students whose first language is other than English, is of critical importance that has yet to be addressed in any systematic fashion across the nation (e.g., King, 1993; LaCelle-Peterson & Rivera, 1994; Linn, 1994; Valdés & Figueroa, 1994; Winograd, Benjamin, & Noll, 2001).

We need to understand the larger issues in education and educational reform, because we often find ourselves torn between the demands for assessment for accountability and the need for assessment to improve instruction. We need to understand why there is such a demand for assessment for accountability and how to deal with that demand in a constructive fashion. We find it useful to think about assessment by identifying what kinds of audience it can serve. For example, assessment can help

- Students become more self-reflective and in control of their own learning.
- Teachers focus their instruction more effectively.
- Educators determine which students are eligible for Title I, programs for the gifted, or special education.
- Parents understand more about their children's progress as learners.

- Administrators understand how groups of students in their schools are progressing.
- Legislators and citizens understand how groups of students across the state or nation are progressing.
- Policy makers and stakeholders monitor the implementation and effectiveness of various reform initiatives, including those that deal with school finance and resource allocation, governance and policy issues, or changes in curriculum.

The best practices in literacy assessment, then, are those that use a variety of appropriate indices to address the needs of different audiences. Thus, the choice does not have to be assessment for accountability versus assessment for instruction. Some states, such as Vermont, Kentucky, and Maryland, have attempted to develop assessments systems that are performance-based, linked to clear standards, support important curricular goals, and are useful for both accountability and instructional purposes. In addition, these assessments are viewed as part of an overall approach to reform that includes professional development, curricular development, and other key changes to the system. As Darling-Hammond (1994, p. 7) notes, the fundamental question “is whether assessment systems will support better teaching and transform schooling for traditional under served students or whether they will merely reify existing inequalities.”

Unfortunately, too many states, districts, and schools continue to use assessments in ways that do reify existing problems and inequalities. The misuse of tests continues despite the large body of research that indicates traditional forms of assessments are based on outdated and inappropriate models of learning; narrow the curriculum in destructive ways; provide results that are misinterpreted and misused; and often produce invalid results that vary widely for individuals and reflect confounded effects related to socioeconomic status, home experiences, or testing conditions (e.g., Darling-Hammond, Ancess, & Falk, 1995; Haney & Madaus, 1989; Winograd, Paris, & Bridge, 1991).

Fortunately, the large body of research on assessment has produced some principles for using assessments wisely. Here are some of our favorites drawn from a number of sources including the National Center for Fair and Open Testing (1998; www.fairtest.org), Fowler and McCallum (1998); Harp (1996); Johnston, 1991; Stiggins (1997); Tierney (1998); and Winograd and Perkins (1996).

The best practices in literacy assessment do the following:

- Focus on important goals and support meaningful student learning.
- Are based on our most current and complete understanding of literacy and children’s development.

- Are based in the classroom rather than imposed from outside.
- Involve students in their own learning and enhance their understanding of their own development.
- Use criteria and standards that are public, so that students, teachers, parents, and others know what is expected.
- Start with what the students currently know.
- Involve teachers (and often students) in the design and use of the assessment.
- Empower teachers to trust their own professional judgments about learners.
- Nourish trust and cooperation between teachers and students.
- Focus on students' strengths rather than just reveal their weaknesses.
- Provide information that is used to advocate for students rather than to penalize them.
- Support meaningful standards based on the understanding that growth and excellence can take many forms.
- Are integral parts of instruction.
- Gather multiple measures over time and in a variety of meaningful contexts.
- Provide educators and others with richer and fairer information about all children, including those who come from linguistically and culturally diverse backgrounds.
- Are part of a systemic approach to improving education that includes strengthening the curriculum, professional development for teachers, and additional support for helping those children who need it.
- Provide information that is clear and useful to students, teachers, parents, and other stakeholders.
- Continually undergo review, revision, and improvement.

The increased use of large-scale, standardized testing for accountability has sparked a large and ever-growing body of literature on the principles for using standardized tests wisely. Given the current political and educational climate, testing for accountability is likely to increase, so the importance of these principles is also likely to increase.

The best single source for appropriate guidance in this area comes from the AERA, APA, and NCME 1999 Standards for Educational and Psychological Testing (AERA, 1999). Here, for example, is AERA's position (based on the 1999 Standards) on the appropriate and fair use of high-stakes testing in pre-K–12 education. This is a long quotation, but it is vitally important that teachers understand these standards given the increased use of high-stakes testing, particularly in the field of reading.

It is the position of the AERA that every high-stakes achievement testing program in education should meet all of the following conditions:

Protection against High-Stakes Decisions Based on a Single Test

Decisions that affect individual students' life chances or educational opportunities should not be made on the basis of test scores alone. Other relevant information should be taken into account to enhance the overall validity of such decisions. As a minimum assurance of fairness, when tests are used as part of making high-stakes decisions for individual students such as promotion to the next grade or high school graduation, students must be afforded multiple opportunities to pass the test. More importantly, when there is credible evidence that a test score may not adequately reflect a student's true proficiency, alternative acceptable means should be provided by which to demonstrate attainment of the tested standards.

Adequate Resources and Opportunity to Learn

When content standards and associated tests are introduced as a reform to change and thereby improve current practice, opportunities to access appropriate materials and retraining consistent with the intended changes should be provided before schools, teachers, or students are sanctioned for failing to meet the new standards. In particular, when testing is used for individual student accountability or certification, students must have had a meaningful opportunity to learn the tested content and cognitive processes. Thus, it must be shown that the tested content has been incorporated into the curriculum, materials, and instruction students are provided before high-stakes consequences are imposed for failing examination.

Validation for Each Separate Intended Use

Tests valid for one use may be invalid for another. Each separate use of a high-stakes test, for individual certification, for school evaluation, for curricular improvement, for increasing student motivation, or for other uses requires a separate evaluation of the strengths and limitations of both the testing program and the test itself.

Full Disclosure of Likely Negative Consequences of High-Stakes Testing Programs

Where credible scientific evidence suggests that a given type of testing program is likely to have negative side effects, test developers and users should make a serious effort to explain these possible effects to policy makers.

Alignment between the Test and the Curriculum

Both the content of the test and the cognitive processes engaged in taking the test should adequately represent the curriculum. High-stakes tests should not be limited to that portion of the relevant curriculum that is easiest to measure. When testing is for school account-

ability or to influence the curriculum, the test should be aligned with the curriculum as set forth in standards documents representing intended goals of instruction. Because high-stakes testing inevitably creates incentives for inappropriate methods of test preparation, multiple test forms should be used or new test forms should be introduced on a regular basis, to avoid a narrowing of the curriculum toward just the content sampled on a particular form.

Validity of Passing Scores and Achievement Levels

When testing programs use specific scores to determine “passing” or to define reporting categories like “proficient,” the validity of these specific scores must be established in addition to demonstrating the representativeness of the test content. To begin with, the purpose and meaning of passing scores or achievement levels must be clearly stated. There is often confusion, for example, among minimum competency levels (traditionally required for grade-to-grade promotion), grade level (traditionally defined as a range of scores around the national average on standardized tests), and “world-class” standards (set at the top of the distribution, anywhere from the 70th to the 99th percentile). Once the purpose is clearly established, sound and appropriate procedures must be followed in setting passing scores or proficiency levels. Finally, validity evidence must be gathered and reported, consistent with the stated purpose.

Opportunities for Meaningful Remediation for Examinees Who Fail High-Stakes Tests

Examinees who fail a high-stakes test should be provided meaningful opportunities for remediation. Remediation should focus on the knowledge and skills the test is intended to address, not just the test performance itself. There should be sufficient time before retaking the test to assure that students have time to remedy any weaknesses discovered.

Appropriate Attention to Language Differences among Examinees

If a student lacks mastery of the language in which a test is given, then that test becomes, in part, a test of language proficiency. Unless a primary purpose of a test is to evaluate language proficiency, it should not be used with students who cannot understand the instructions or the language of the test itself. If English language learners are tested in English, their performance should be interpreted in the light of their language proficiency. Special accommodations for English language learners may be necessary to obtain valid scores.

Appropriate Attention to Students with Disabilities

In testing individuals with disabilities, steps should be taken to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement.

Careful Adherence to Explicit Rules for Determining Which Students Are to Be Tested

When schools, districts, or other administrative units are compared to one another or when changes in scores are tracked over time, there must be explicit policies specifying which students are to be tested and under what circumstances students may be exempted from testing. Such policies must be uniformly enforced to assure the validity of score comparisons. In addition, reporting of test score results should accurately portray the percentage of students exempted.

Sufficient Reliability for Each Intended Use

Reliability refers to the accuracy or precision of test scores. It must be shown that scores reported for individuals or for schools are sufficiently accurate to support each intended interpretation. Accuracy should be examined for the scores actually used. For example, information about the reliability of raw scores may not adequately describe the accuracy of percentiles; information about the reliability of school means may be insufficient if scores for subgroups are also used in reaching decisions about schools.

Ongoing Evaluation of Intended and Unintended Effects of High-Stakes Testing

With any high-stakes testing program, ongoing evaluation of both intended and unintended consequences is essential. In most cases, the governmental body that mandates the test should also provide resources for a continuing program of research and for dissemination of research findings concerning both the positive and the negative effects of the testing program. (AERA, 1999, adopted July 2000; www.aera.net/about/policy/stakes.htm)

These principles are important for classroom teachers to understand. Often, the best practices in literacy assessment mean being an informed and effective advocate for the fair, limited, and reasonable use of tests in today's schools.

RESEARCH-BASED PRACTICE IN CLASSROOM ASSESSMENT

The term “research-based practice” covers a great deal of territory, including everything from high-stakes standardized testing to a range of classroom-focused evaluation strategies. For example, at the national level, using the NAEP to assess the reading achievement of large groups of students is based on solid psychometric research. These scores are used by policy makers to evaluate the effects of large-scale educational initiatives. At a more local level, individual teachers use a variety of research-based practices to learn about their particular students and their own teaching,