ED 356 457   CS 011 269

AUTHOR   Pearson, P. David; Stallman, Anne C.
TITLE   Approaches to the Future of Reading Assessment:
                Resistance, Complacency, Reform. Technical Report No.
                575.
INSTITUTION   Center for the Study of Reading, Urbana, IL.
PUB DATE   May 93
NOTE   17p.
PUB TYPE   Viewpoints (Opinion/Position Papers, Essays, etc.)
                (120) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   Educational Change; Educational History; Elementary
                Secondary Education; Higher Education; Reading
                Ability; *Reading Achievement; *Reading Tests;
                Standardized Tests; *Student Evaluation
IDENTIFIERS   *Alternative Assessment; *Authentic Assessment;
                Educational Issues

ABSTRACT
        Noting that assessment plays a prominent role in all
levels of the American educational system, this report characterizes
the realities of reading assessment and its historical evolution. The
report then characterizes possible approaches to shaping the future
of reading assessment--resistance, complacency, and reform. The
report considers the rationale, feasibility, and possible
consequences of each approach. The report concludes that: the
realities of educational politics in the United States mean that
wide-scale, efficient forms of assessment will continue to be needed,
or at least mandated by policy makers; and that only through reform
can assessment become a positive force in the lives of children and
teachers. (Contains 27 references.) (Author/RS)

ED356457

Center for the Study of Reading

TECHNICAL
REPORTS

BEST COPY AVAILABLE

CS011269

ERIC
Full Text Provided by ERIC

CENTER FOR THE STUDY OF READING

Technical Report No. 575

APPROACHES TO THE FUTURE OF
READING ASSESSMENT: RESISTANCE,
COMPLACENCY, REFORM

P. David Pearson
Anne C. Stallman
University of Illinois at Urbana-Champaign

May 1993

1992-93 Editorial Advisory Board

Abstract

Assessment plays a prominent role in all levels of the American education system. This report characterizes the current realities of reading assessment and its historical evolution. It then characterizes possible approaches to shaping the future of reading assessment--resistance, complacency, and reform. The rationale, feasibility, and possible consequences of each approach are considered.

APPROACHES TO THE FUTURE OF READING ASSESSMENT:
RESISTANCE, COMPLACENCY, REFORM

Assessment figures prominently at all levels of the education system, from the individual student in the classroom to the policy maker in the statehouse. In this report, we first characterize the current realities of reading assessment, then explain how these realities evolved. Finally, we offer suggestions for reading educators to consider as they come to grips with assessment's role in shaping the range of students' activities, the quality of teachers' instructional decisions, and the nature of school curricula.

The Current Realities of Assessment

Despite the recent and well-publicized calls for performance-based assessment (Wiggins, 1991), standardized, commercially produced paper and pencil tests remain the most prevalent form of assessment in the United States. Continuing a tradition that dates to the 1920s, these tests rely almost exclusively on multiple-choice, single-correct-answer formats. They may also contain relatively short, information-laden story snippets--often about an obscure topic--which are followed by a set of questions requiring students to draw inferences from the most obscure facts in the stories. Such tests pervade the education system: basal reading program tests tend to be closely modeled on them (although recent editions show some evidence of change), and even in kindergarten testing, multiple-choice items predominate (Stallman & Pearson, 1990). If one considers the various forms of formal and informal tests that a student is likely to encounter from kindergarten through high school, it is well within the realm of possibility that he will complete over 20,000 test items in a school career, or an average of about 46 items per week.

Standardized testing maintains its prominent position in the nation's schools because of both economic and political reasons. For example, as Americans we want our children to receive the best education possible, but we also want to get the most from the education dollars we spend. As a result, we have developed an "accountability mentality" for evaluating the education system, and opinion polls indicate consistently that the general public believes that standardized test scores are the best indicators of the quality of education children are receiving (Elam, 1990; Elam & Gallup, 1989). The economic force exerted by these scores is evident by uses made of them. In some areas, for example, newspapers publish schools' test scores. Realtors then may quote the scores to prospective property buyers as part of their sales pitch, and, as a result, property values rise and fall according to the quality of the schools, as indexed by test scores (Pearson et al., 1990).

Test scores can take on a political edge when state legislators cite them as the criteria for deciding how special funds will be allocated to schools and school districts.

These economic and political dimensions of testing have serious curricular consequences, causing some educators to assign ever-increasing importance to testing. Some educators make curricular and instructional decisions for the *sole* purpose of improving test scores. Consequently, there is a growing body of evidence that, in many cases, assessment is affecting instruction, both directly and indirectly.

Assessment *directly* affects instruction when educators change their instructional practices for the explicit purpose of improving test scores. The most common example of this is when teachers and students take time from regular instruction to prepare for a particular test. This practice has been well-documented in states such as Florida, Texas, and North Carolina, where a great deal of emphasis is placed on assessment, and it continues, not because it produces better educated students, but because it produces higher test scores. This begins the game that we call "high-stakes assessment." Schools with higher test scores gain prestige over those schools with lower scores; instructional programs that demonstrate greater gains are more likely to receive increased financial support; students with higher scores on tests

such as the SAT get into better colleges. And so tests and test scores have immediate and direct consequences for the entire education community--districts, schools, teachers, and students.

Assessment *indirectly* affects instruction in a variety of ways. For example, a district may decide to adopt a particular basal reading program on the belief that the new program will somehow raise students' reading test scores. A school may shift the focus of its curriculum to align it more closely with what is covered on tests. These effects stem from the faith that school boards, administrators, and teachers put in test makers. They believe that test makers have some special knowledge of what needs to be taught, and that the tests they create reflect that knowledge. This belief is reinforced by education publishers, who market a particular program by stressing the close ties between its content and what's tested on the most widely used tests. A corollary of this belief in test makers is that if they don't include something on their tests, it must not be important and therefore should not be taught. And so tests have indirect effects on curricula--on both what is and is not taught.

In response to this situation, some educators have argued that because assessment appears to drive instruction, tests should be changed so that when teachers do teach to them, they will be teaching students the kinds of skills and knowledge that really matter (e.g., Resnick & Resnick, 1990; Wiggins, 1991). Proponents of this argument often call for the use of performance-based forms of assessment in place of standardized tests.

Advocates of standardized tests counter by arguing that performance-based assessment is simply too expensive for education systems that suffer chronically from a lack of financial support. They further argue that standardized tests have many strengths. Unlike time-consuming alternative assessments, they argue, standardized tests are efficient and take very little time away from instruction. They also maintain that standardized tests are more fair than performance-based assessments because they are more objective, free from the biases of any individual, and culturally neutral. Their use, proponents maintain, ensures that everyone will have the same chance for success; that is, the tests provide a "level playing field" for everyone, regardless of income, cultural background, or educational experience.

Thus although standardized tests have been the target of much criticism in the past decade, they nonetheless continue to exert considerable influence.

The Evolution of Assessment

To understand current assessment practices, it is necessary to look to their beginnings in the early years of this century. Edward L. Thorndike's handwriting scale, published around 1914, is considered to be the first standardized test, and it set the stage for the kind of assessment that flourished in the years following. In creating his scale, Thorndike measured the performance on handwriting tasks of thousands of students. He then used their performance scores to establish norms (average levels of performance) for students of various ages. These norms became standards that teachers could use to evaluate the progress of their own students. Our current notion of a "norm-referenced" test stems from these early efforts by Thorndike and others to establish educational standards with reference to the average levels of performance of normal populations of students.

Several factors contributed to an favorable environment for the growth and popularity of the standardized testing movement: the beginning of World War I, the introduction of compulsory education, the growth of the civil service system, and the rise of scientific objectivity as a driving force in the social sciences.

World War I's contribution to the movement came about because of the perceived complexities of "modern warfare." It was deemed important to find ways to determine which recruits had the skills to

cope with these complexities. Not surprisingly, norm-referenced, multiple-choice tests of reading and writing skills were adopted as economical and efficient screening mechanisms.

The spread of compulsory education created the need for ways to gather objective evidence to determine which students should be advanced to the next grade and who would get into or out of certain institutions or programs--what is now called gatekeeping. One of the earlier examples of this gatekeeping function occurred with our youngest children. Reading readiness tests were developed in the 1920s in response to educators' concerns about children who were unsuccessful in first grade, usually because they lacked "prereading" skills. Readiness tests were used to determine which children could and should receive regular reading instruction. For those children who were not ready, the tests also served a diagnostic function--to pinpoint the specific skills these children lacked so that teachers could provide appropriate instruction (Stallman & Pearson, 1990).

The most important spur to the standardized testing movement came not from educators, however, but from bureaucrats in charge of the new civil service systems run by federal, state, and local governments. Reform-minded civil service bureaucrats sought ways to determine a worker's capabilities without relying upon the interviews and work samples that had been used for years as thinly disguised tools for denying jobs to members of certain racial and ethnic groups. In standardized tests, these bureaucrats found the objectivity and neutrality they were seeking; they had a tool, they thought, that was color-blind, culture-free, and fair to everyone seeking employment.

The appeal of "objective" testing also reflected the growing influence of the scientific objectivity movement. Because standardized tests were supposed to be based on scientific principles, they were considered free of bias and, as such, inherently fairer to everyone. In short, they were regarded as reliable instruments that could be counted on to yield consistent, objective results over repeated administrations and in a variety of contexts. While all of these factors helped to create a favorable environment for the development of standardized testing, it was a technological innovation that secured its place in American life--the invention of the IBM 805 scanner in 1934. This device cut the cost of scoring tests to one tenth of what it had been and increased the appeal of their use dramatically. Now, not only were the tests objective, they were efficient to administer and score (Pearson & Dunning, 1985; Resnick, 1982).

The 1960s saw two watershed developments in standardized testing. The first of these developments grew out of the passage of the Elementary and Secondary Education Act of 1965, which provided for Title I, now Chapter 1, programs. The reauthorization of this act in the late 1960s contained an important provision: States had to agree to be accountable for how well the allocated federal funds were spent. By agreeing to test the progress of Chapter 1 students, a major precedent was set. In effect, educators implicitly accepted the premise that student improvement--measured by standardized tests--was the natural outcome of compensatory education. That premise now applies not only to compensatory education but to education in general.

The second watershed development came in the late 1960s, when instruction was linked to assessment in another significant way. Prompted by the egalitarian appeal of concepts such as mastery learning (Bloom, 1968; Carroll, 1963), a whole new approach to assessment was developed (Pearson & Dunning, 1985). Instead of holding instruction constant--by teaching all students in the same way--and allowing achievement to vary, proponents of mastery learning advised teachers to hold achievement constant and allow instruction to vary on a number of dimensions: (a) student time on task (how long students take to achieve mastery), (b) teacher time on task (how much extra instruction teachers provided), and (c) skill prerequisites.

The idea that achievement should be held constant was inconsistent with the philosophy behind norm-referenced tests, which reference achievement to a relative standard (the mean of the distribution) rather than some absolute criterion of achievement. One outgrowth of the mastery learning movement

was the development of criterion-referenced tests, which were designed to hold achievement constant by requiring students to achieve a prespecified, absolute level of mastery (usually 80% correct) on some target behavior. The mastery notion has had a significant and long-lasting, if unintended, influence on reading assessment. When mastery learning ideas were introduced into the reading field during the 1960s, task-analytic traditions dominated instructional design (e.g., Gagne). Applied to reading, this suggested that the complex task of reading be broken down into component parts so that students could deal with each component separately. Thus, the notion of mastery learning was applied not to *reading itself* as a real-world process but to the *component parts* of reading. Students were tested on their mastery of the subskills of reading such as consonant blends, vowel digraphs, sequencing, and locating the main idea. This led to the powerful assessment practice of skills-management systems. First promoted as free-standing entities, such as the Wisconsin Design, these systems soon became infused into basal reading programs.

In fairness, it must be noted that nothing in Carroll or Bloom's conceptualizations of school learning dictated that mastery notions had to be applied in this way; the notion of mastery could just as easily have been applied to reading as a process. However, the instructional paradigm of the time mandated that the notion of mastery had to be applied at the subskill level. The decision to do so had an incredible impact on the development of reading curriculum and assessment materials during the next two decades.

The most recent important development in standardized testing came in the early 1980s--the development of what is now called outcomes-based education. The basic idea behind outcomes-based education is that schools ought to be willing to hold themselves accountable to specific performance objectives, such as "the mean score of students in District X on Test Y will equal or exceed the 60th percentile." This represents a very different linkage between assessment and instruction than the linkage which stemmed from mastery learning. In the skills management systems that emerged from mastery learning in the late 1960s, assessment and instruction was linked at the level of subskills; within the framework of outcomes-based education, this link was made at a more global level. Spurred by the call to action made by *A Nation at Risk*, and by the apparent success of the outcomes-based programs adopted by states such as Florida, Texas, South Carolina, and Maryland (Popham, Cruse, Rankin, Sandifer, & Williams, 1985), many concluded that teaching to the test was not such a bad practice. If the test scores in a state could be increased significantly by familiarizing teachers with the specifics of what was to be tested, it seemed reasonable to have them teaching to tests that covered what children needed to know.

During the last few years, however, educators have come to realize that teaching to these tests often has the effect of narrowing the curriculum to those outcomes that are easily measurable (Shepard & Dougherty, 1991). Koretz, Linn, Dunbar, and Shepard (1991) have dramatically illustrated this phenomenon. They have worked with a large school district that has assumed a high-stakes assessment profile. Scores there are a matter of public record, and there are consequences for low scores. In 1986, the district used Standardized Test A. In 1987, when it switched to Standardized Test B, average scores, using the norms provided by the new tests, dropped over a half grade level in comparison to the previous year. But then an interesting development occurred. In each year from 1987 to 1990, the average scores, computed at the school level, rose substantially. By 1990, for example, scores for the district's third graders were almost a grade (in grade-norm units) higher than they were in 1987, the first year of the new test. In 1990, the district gave a second test: the old Standardized Test A. Scores on this test were half a grade lower than on the 1990 Test B. (However, this is not really a fair comparison because of the differences in norming populations between Tests A and B.) The 1990 Test A scores were also compared to the 1986 Test A scores. The 1990 scores had dropped a full grade level. These findings strongly suggest that the improved scores on Standardized Test B between 1987-1990 were due to teaching to the test.

The researchers made one other comparison. In 1990, they gave alternative tests in both math and reading. The items on the alternative tests were more like everyday classroom assignments (e.g., solving math problems, writing in response to reading, responding to items with multiple correct answers). They also gave these alternative tests, along with Standardized Test B, in a district that, while demographically similar to the first test district, did not assume a high-stakes profile. They compared the performance between these districts on both types of tests. They found that students in the first district performed just as well as those in the second district on the standardized tests, but scored consistently scored below them on the alternative assessments. Again, a conclusion that is hard to avoid is that teaching to the test narrows the curriculum.

Next Steps

Given the current realities of assessment and the historical forces that have brought us to those realities, we see three possible approaches to shaping the future of reading assessment. First, educators and researchers can oppose formal assessments altogether and return to the subjective tradition of a century ago; an approach we characterize as *resistance*. Second, they can learn to live within the parameters of the current situation; an approach we label *complacency*. Or, third, they can try to improve assessment by changing the system from within; an approach we call *reform*. We will consider the rationale for, the feasibility of, and the possible consequences of each of these approaches.

Resistance. There are many powerful arguments for resisting the use of formal assessments as we presently know them--that is, standardized tests. The major argument in favor of resistance is one we described earlier: teaching to the test has the effect of narrowing the curriculum.

A second argument for resistance is that when educators use tests to shape instruction, they essentially invalidate test results. Most standardized tests attempt to generalize beyond the sample of items actually found on the test and are predicated on the assumption that educators will *not* teach to them. Classroom realities, however, all too often reveal this to be a false assumption. Imagine, for example, that the test to be given to your class has 40 vocabulary items. Suppose that a worksheet suddenly appears in your mailbox with vocabulary exercises that allow your students to practice those 40 words-- and that the mean vocabulary score subsequently jumps from the 47th to the 75th percentile. Do these results lead to the conclusion that your students have a much richer knowledge of vocabulary than they had earlier? Of course not. The students' scores have risen, not because their vocabularies have been developed, but because they have inside information about the test.

A third argument for opposing formal assessments rests upon criticism of what the current tests are testing and how they are doing it. Over the past 30 years, there have been many changes in the way in which reading is conceptualized and defined; standardized tests, however, have remained relatively impervious to these new concepts and definitions. Reading is now generally thought of as an interactive process in which the reader uses available resources such as background knowledge, the text, and the context of the situation to construct meaning. Standardized tests, however, continue to be based on the notion that reading is an aggregation of subskills. This mismatch has spawned several criticisms of standardized tests:

1. Standardized tests measure something by using short passages and multiple-choice formats, but what they measure is not what most people now understand reading to entail. Reading is a generative process; moment by moment, readers build and rebuild models of meaning with the thoughts, ideas, and images that come to them. Filling in bubbles on a scannable answer sheet does not capture the dynamic quality of this process.

2. By using so many different short passages, standardized tests mask the inherent, and important, relationship between background knowledge and comprehension. The practice of using short passages

more or less guarantees that students with good general verbal ability will do best on standardized tests. Again, test designers need to acknowledge that reading ability is a dynamic, not a static or a general, phenomenon. Different variables, such as student interest and background knowledge, will differentially affect a student's performance.

3. Standardized tests do not contain enough inferential and critical reading items. As a consequence, they provide only a partial picture of what students can do in response to reading passages. Furthermore, as we have already noted, teaching to such trivializing tests will result in simplistic and narrowly focused curricula.

4. Although standardized tests were frequently touted as culturally neutral, contemporary evaluation of these tests has revealed how they are biased against racial and linguistic minorities. Even if the more obvious forms of cultural bias, such as topic selections and ethnic representations, are discounted, these tests are biased in other ways. For example, the performance of Latino students, even those who speak English relatively well, suffers on these tests because of the extra language processing that these students must engage in when reading English. The precise time limits of standardized tests disadvantage them. When given tests in an untimed context, however, their performance increases substantially (García, 1991).

5. The curriculum-narrowing effect of standardized tests is exacerbated for racial and linguistic minorities. Dorr-Bremme and Herman (1986), for example, found that the curriculum of low-income students was influenced by commercial tests to a much greater degree than was true for more affluent children. Low-income students are more likely to be in programs funded by federal entitlement funds, for which commercial assessment is mandatory. Teachers and administrators responsible for these programs understandably want their students to demonstrate progress. However, because progress is often defined in terms of students' performance on tests of subskills, low-income students are likely to spend their time and energy practicing such skills in the mistaken belief that this will lead them to score higher on the tests (see García & Pearson, 1991).

Taken together, these arguments provide strong evidence that formal assessment as we know it should be eliminated. Before this can be accomplished, however, the demand for accountability means that some other system for evaluating educational practice acceptable to the public must be found. This new system must assure the public that their children are getting a good education and that the money being spent on education is being used effectively. Any new system of accountability must also address the public's concern about fairness and objectivity. We believe that these concerns and demands make it difficult to imagine the complete elimination of standardized testing in the immediate future. The continuing reliance on tests leads us, in turn, to consider the two other approaches to assessment.

Complacency. The second approach to assessment in the future is complacency. This approach is predicated on the belief that because educators cannot change the political climate and eliminate formal assessments, they should just learn to live with them. Some advocates of this approach argue, in fact, that teaching to tests is not so bad because it forces educators and the educational system to be accountable, even if it does so imperfectly.

Complacency is, probably by default, the approach taken by most educators today. Their reasons for doing so are many and varied. Tests do have political value and they have society's general approval. Parents argue that because they had to take tests, their children should have to too, and many educators and politicians view them as an indispensable tool in evaluating program effectiveness.

The net result is a strong rationale for maintaining the status quo. However, there are those who believe that changes need to be made if the present system is to remain credible and that working toward change is a worthwhile enterprise.

Reform.   Reformers acknowledge that formal assessment is deeply entrenched in the American educational system. They believe, however, that changes in test design and in how test results are interpreted and used will make the assessment situation more acceptable.

One means of reforming assessment is to make sure that tests cover what is essential for students to know and be able to do. Some states have adopted this approach with their state reading assessments. Illinois, Michigan, Maryland, Wisconsin, and California, for example, have designed assessments on the assumption that a well-designed test can have a positive influence on school curricula. The basic idea is that if the tests cover what really matters, then teaching to the tests might actually be beneficial. An example of these reforms comes from Illinois, where the state reading assessment developed in the late 1980s was based on the premise that reading is an integrated, constructive, empowering, and meaning-based process. Those of us who helped develop the Illinois assessment anticipated that if schools developed reading programs in which students spent a lot of time reading, writing, thinking, and evaluating, the students would do just fine on the state tests. Our hopes were borne out; indeed, many districts in Illinois have used the state assessment to convince local educational policy makers of the need to revise school curricula.

In other instances, horror stories testify to the fact that the classroom implementation of reform ideas can be undermined and perverted, either intentionally or unintentionally. As part of the statewide assessment of reading in Illinois, for example, we attempted to measure students' knowledge of the topics of the passages they read as a way to explain their resulting comprehension scores. Had educators responded to this feature of the assessment by developing a range of activities to help students see the connection between what they know and what they read, the outcome would have been positive and appropriate. However, in some school districts, educators prepared students for this part of the assessment by requiring them to complete prior-knowledge worksheets for *every* story they read in their basal reading program. In these districts, what we witnessed was an impulsive response to a surface feature of the test rather than a thoughtful response to the principles upon which the test was built.

In addition, we used a structural device, the story map, to develop questions that gave students opportunities to exhibit an integrated understanding of a passage. Under the apparent assumption that if story maps are good for test writers, they must also be good for students, educators in some districts required students to write story map summaries of every story they read.

Stories such as these underscore the fact that simply changing tests is, at best, only a partial solution to the current assessment problems. It may be in the nature of such high-stakes assessment that educators will feel pressured to ensure good student performance on the tests and respond to that pressure by overreacting to surface features of the assessment processes rather than their underlying principles.

In recent years, there have been numerous calls for alternative forms of assessment to realign the links between instruction and assessment (Valencia, 1990). Whether they use terms such as informal assessment, portfolio assessment, performance-based assessment, teacher-developed assessment, or student self-assessment, these calls for more authentic forms of assessment have in common a healthy distaste for standardized tests and a conviction that if truly liberating standards of performance--and tests to go with them--can be developed, liberating curricula will follow.

Proponents of authentic assessment argue that the real purpose of classroom assessment is to help teachers make informed instructional decisions (Wixson, 1991) and to inform both teachers and students about achievement and progress in school (Tierney, Carter, & Desai, 1991). Arguments for authentic assessment have at their center the beliefs that if assessments are to be meaningful and useful (a) they must reflect current knowledge about the dynamics of reading; (b) they should yield information about performance on real, not contrived tasks; (c) they should be a natural part of classroom activity, not an intrusion upon it; and (d) they should be viewed as beneficial to both teachers and students as they

assume joint responsibility for the educational process (Au, Scheu, & Kawakami, 1990; Lipson, 1990; Peters, 1991; Tierney et al., 1991).

Two current assessment efforts typify this version of reform. The New Standards Project (Resnick & Resnick, 1990) has the explicit goal of developing a voluntary national assessment for individual children based on the same logic that underlies many of the new state reading assessments. Project members contend that a national test that measures what matters will spur the development of the kind of curricula that will allow the United States to recapture its position of world economic leadership. It was in response to the calls for reform that developers of the 1992 National Assessment of Educational Progress (NAEP) in reading included an experimental component in which portfolios, individual interviews, and samples of students reading books orally were included in the assessment.

One potential application of the alternative assessment agenda deserves special mention: the call to replace standardized tests with classroom-based assessments. The logic behind this particular proposal runs as follows: What if standardized tests were eliminated in an entire school district? What if they were replaced with some kind of portfolio approach, one in which district administrators, teachers, and students could all have a voice in specifying what would be included in the portfolio? Most educators would agree that such an approach would provide valuable documentation about the progress of individual students, but they are unsure about what it would do to program evaluation at the school and district levels. Is there some way to translate portfolio entries into data for wide-scale evaluation? Perhaps. One way was found by teachers and administrators in a suburban Chicago district, who had convinced the school board to adopt just such an approach. In this district, teachers collect portfolio entries and score them for their own classroom and for individual students. Then, to obtain school- and district-level data, outside auditors are brought in to take a random sample of portfolios and of entries from these portfolios. These samples are then scored independently and the results are tabulated by school and for the district. The school and district reports contain the kinds of tables and charts that can be constructed from standardized tests, with the obvious omission of comparative state or national norms. However, teachers, administrators, and the local school board can still evaluate what it means, for example, when 74% of the students achieve a competent standard on a writing-in-response-to-reading task. They can also compare the 1992 fifth-grade class with the 1991 class on this, and just about any other criterion. The advantage of using this assessment system is that there are not artificially determined factors present to drive instruction down counterproductive curricular routes. The same data that guide a teacher's decisions within the classroom also provide information that can be used for larger accountability purposes.

While the premise that underlies alternative assessment seems reasonable, several issues still need to be resolved if these new types of measures are to be the future of assessment. The main issue is establishing credibility in the mind of the public: Can the information produced by these alternative measures be trusted? Are they fair to all students, or will they discriminate against particular groups? The importance of establishing credibility with the public cannot be underestimated -- for example, an attempt to include a writing sample in the latest version of the SAT was thwarted by an advisory board member who was convinced that essays were biased against Asian students.

Another issue that must be addressed centers on the utility of alternative measures. While they may be very useful for making decisions at the classroom level, can they really be used for making decisions at the larger district, regional, state, or national level? The one successful attempt to do so that we noted is encouraging, but it does not ensure that citizens and policy makers in other settings accept such an approach. Finally, there is the issue of what standards are to be used to evaluate alternative assessments. It is not clear, for example, that the conventional standards of reliability, validity, utility, efficiency, and objectivity are applicable to alternative measures.

Conclusion

The current realities of educational politics in the United States means that wide-scale, efficient forms of assessment will continue to be needed, or at least mandated by policy makers. Improving the way we assess our students' reading ability, therefore, seems to be a reasonable course of action. Because tests appear to have such a powerful influence on curriculum, we must find a way to come to terms with them--to resist them, accept them, or reform them. Our preference is for reform. And we believe that the development, validation, and evaluation of new performance-based assessments is the most appropriate reform strategy. To remain complacent about our current tests is to doom many children to narrow curricula. To simply resist their use is to ignore their influences, both negative and positive. Our only real option is reforming them, so that they reflect what we know about reading and learning. Only through reform can assessment become a positive force in the lives of children and teachers.

References

Au, K. H., Scheu, J. A., & Kawakami, A. J. (1990). Assessment of students' ownership of literacy. *The Reading Teacher, 44,* 154-156.

Bloom, B. S. (1968). *Learning for mastery. Evaluation comment.* Los Angeles: University of California, Center for the Study of Evaluation.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723-733.

Dorr-Bremme, D. W., & Herman, J. L. (1986). *Assessing student achievement: A profile of classroom practices.* Los Angeles: University of California, Center for the Study of Evaluation.

Elam, S. M. (1990). The 22nd annual Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 72,* 41-55.

Elam, S. M., & Gallup, A. M. (1989). The 21st annual Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 71,* 41-54.

García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26,* 371-392.

García, G. E., & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. H. Hiebert (Ed.), *Literacy in a diverse society: Perspectives, practices, and policies* (pp. 253-278). New York: Teachers College Press.

Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalizations across tests.* Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Lipson, M. Y. (1990). Evaluating the reading context. *The Reading Teacher, 44,* 330-332.

Pearson, P. D., & Dunning, D. B. (1985). The impact of assessment on reading instruction. *Illinois Reading Council Journal, 13,* 18-29.

Pearson, P. D., Stephens, D., Commeyras, M., Roe, M., Shelton, J., Stallman, A. C., & Scharer, P. (1990, May). *Assessment and decision-making in the schools.* Paper presented at the annual meeting of the International Reading Association, Atlanta.

Peters, C. W. (1991). You can't have authentic assessment without authentic content. *The Reading Teacher, 44,* 590-591.

Popham, W. J., Cruse, K. L., Rankin, S. C., Sandifer, P. D., & Williams, P. L. (1985). Measurement driven instruction: It's on the road. *Phi Delta Kappan, 55,* 628-634.

Resnick, D. P. (1982). History of educational testing. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies (Part 2).* Washington, DC: National Academy Press.

Resnick, L. B., & Resnick, D. P. (1990). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction*. Boston: Kluwer.

Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction*. Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Stallman, A. C., & Pearson, P. D. (1990). *Formal measures of early literacy* (Tech. Rep. No. 511). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwell, MA: Christopher-Gordon.

Valencia, S. W. (1990). Alternative assessment: Separating the wheat from the chaff. *The Reading Teacher, 44*, 60-61.

Wiggins, G. (1991). Standards, not standardization: Evoking quality student work. *Educational Leadership, 48*, 18-25.

Wixson, K. K. (1991). Diagnostic teaching. *The Reading Teacher, 44*, 420-422.

Author Note