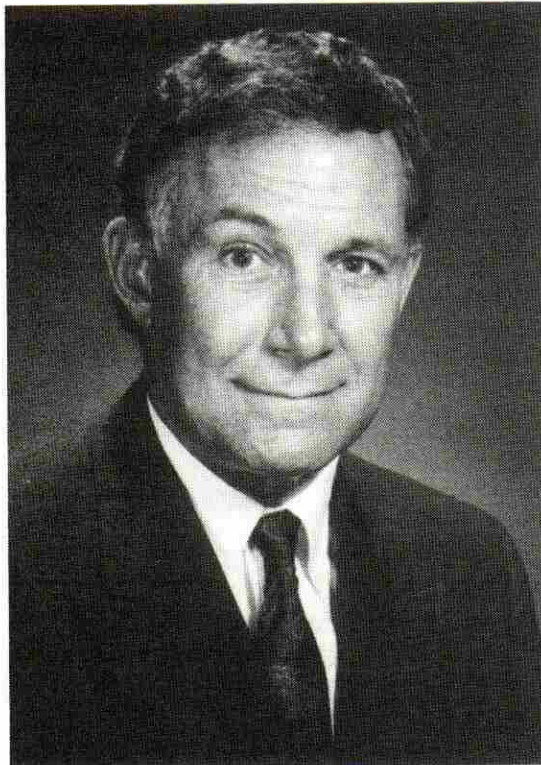


Roger Farr

Putting it all together: Solving the reading assessment puzzle

Farr, a Professor of Education and Director of the Center for Reading and Language Studies at Indiana University, has written widely on the topic of reading assessment.



Reading assessment has become a genuine puzzle. Confusion and debate continue about what the goals of school assessment of reading should be and about what types of tests and other assessments are needed to achieve those goals. That debate should focus on the purposes for assessment and whether current tests achieve those purposes. Too often, however, the focus of the debate is on the latest testing panacea. In this article, I first examine the complex components of the assessment puzzle. Next I propose a solution to the puzzle that involves linkages among various assessment audiences and approaches. I conclude with a few remarks about how school districts in the United States might pull together all the pieces and solve the assessment puzzle for themselves.

Examining the pieces of the assessment puzzle

The pieces of the puzzle represent many types of assessments, critical attitudes about them, and attempts to challenge or improve them. One of the truly puzzling aspects of reading assessment to many educators is that the amount of testing appears to increase at the same time that criticism of it intensifies

The "Distinguished Educator" series provides prominent literacy educators an opportunity to present their personal views on topics about which they are recognized experts.

(Farr & Carey, 1986; McClellan, 1988; Salganik, 1985; Valencia & Pearson, 1987).

Criticism of schools has led to more assessment

Public disappointment with student achievement has led to extensive criticism of U.S. schools. This disapproval intensified in the 1950s with a focus on reading. Reading assessment conducted to prove or disprove the criticism has received a great deal of attention ever since. Could Johnny read or not, and how well or how poorly? By the 1960s, and beyond, score declines on tests used to predict how well high schoolers would do in college compounded public concern and criticism (The National Commission on Excellence in Education, 1983).

The conviction that many students were receiving high school diplomas and yet were almost totally illiterate became firmly established in the public's mind (Purves & Niles, 1984). The Peter Doe case in California exemplified that concern (Saretsky, 1973). The case concerned a high school student who sued the school district for graduating him without teaching him to read. As a result of this kind of dissatisfaction with educational outcomes, the use of standardized, norm-referenced assessment intensified, and state minimum competency testing programs proliferated (Madaus, 1985; Salmon-Cox, 1981).

The data to determine whether scores on reading tests were deteriorating over time is sketchy at best and tends not to substantiate dramatic declines in the reading performance of U.S. students over the years (Farr & Fay, 1982; Farr, Fay, Myers, & Ginsberg, 1987; Stedman & Kaestle, 1987). Nonetheless, the public has remained convinced that performance has dropped rather dramatically. Further, the prevalence of minimum competency programs has not significantly altered the conviction of the public and press that student achievement, particularly in reading, continues to deteriorate.

This unabated critical concern was at least partly responsible for the establishment of the National Assessment of Educational Progress (NAEP), an ongoing federally mandated study that now provides some reading performance data over time. Any declines it has depicted are small compared to the pub-

lic's determined assumptions (Mullis, Owen, & Phillips, 1990). And although careful analyses of the ACT and SAT score declines has cited several reasonable causes other than poor schools, that phenomenon did much to sustain and cement public conviction and the demand for accountability testing (Popham, 1987; Resnick, 1982).

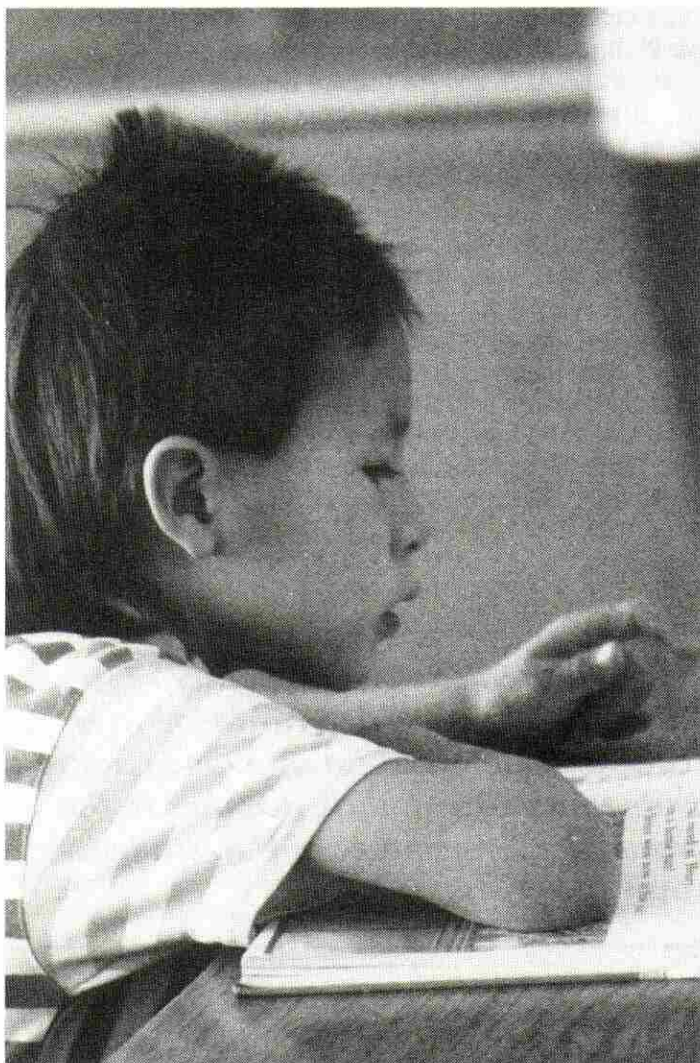
The continuing debate about the quality of U.S. schools has now given rise to a new focus on standards and assessment. At the same time that they reaffirm their conviction that children are not learning in school, critics like Chester Finn (1992) echo the call from the White House "for new American achievement tests" that compare student performance

The data to determine whether scores on reading tests were deteriorating over time is sketchy at best and tends not to substantiate dramatic declines in the reading performance of U.S. students over the years.

to "world class standards" that would be set as criterion references. President Bush (1991) has called for "voluntary national tests for 4th, 8th, and 12[th] graders in the five core subjects" to "tell parents and educators, politicians and employers, just how well our schools are doing."

The search for alternative assessments has also led to more assessment

In addition to dissatisfaction with the schools, there has been a quest for assessments that are closely aligned with more holistic views of language development. Some curriculum theorists concerned with the mismatch between curriculum and assessment have determined that if curriculum is to change, the reading tests must change. This has brought about a proliferation of new assessments—both formal and informal (Brown, 1986; Burstall, 1986; Priestley,



"Students need to become good self-assessors if they are to improve their literacy skills. They need to select, review, and think about the reading and writing they are doing." Photo by Jan McKean, Little Bighorn Photos

1982; Stiggins, Conklin, & Bridgeford, 1986).

Included in this mix have been modifications of conventional tests with new item formats and the addition of the assessment of behaviors not often included on traditional tests, such as background knowledge, student interests and attitudes, and metacognition. Other assessments in reading have taken an entirely different approach to assessment, relying entirely on student work samples collected in portfolios (Jongsma, 1989; Valencia, 1990; Wolf, 1989). Portfolios have

themselves taken many different forms from *show portfolios*, which include only a few carefully selected samples, to *working portfolios*, which include a broad sample of work and which are used to guide and organize daily instruction. In addition, numerous professional publications have published articles calling for the use of a broader range of teacher observations and informal assessment techniques (Cambourne & Turbill, 1990; Goodman, 1991).

Different audiences need different information

Thus, it seems that the increased amount of testing has resulted from greater accountability demands as well as from attempts to find alternatives to traditional assessments. In order to bring some sense to this proliferation of assessment, we need to understand that tests have only one general purpose: Tests should be considered as nothing more than attempts to systematically gather information. The information is used to help children learn about their own literacy development and to give teachers and others concerned with students' literacy the information they need for curriculum planning. *The bottom line in selecting and using any assessment should be whether it helps students.*

A book that I first authored more than 20 years ago regarding the assessment of reading was entitled *Reading: What Can Be Measured?* (Farr, 1970; Farr & Carey, 1986). I have always felt that the title gave the wrong focus to the review of assessment issues. That book should have been entitled, *Reading: Why Should It Be Measured?* We need to consider who needs information about reading, what kind of information is needed, and when it is needed. Only then can we begin to plan for more sensible assessment.

In order to think more clearly about overall assessment plans, we need to know why we want to test. There are, of course, different groups that need information. Without considering these groups and their information needs, the assessment program in any school system will remain as a set of jumbled puzzle pieces. The general distinctions between audiences are covered in Figure 1.

The public. Members of the general public, who make decisions through their elected officials, including school boards, have a

vested interest in the future of children and in their effective and cost-efficient instruction. It is recognized as vital to Americans' and their nation's future that schools produce educated students. Indeed, the most recent federally supported efforts to improve education have been on establishing standards that presumably will result in the development of assessments related to those standards. At the present time, those involved with establishing the standards are moving in the direction of holistic kinds of performance assessment.

Administrators. Ideally school administrators would rely most heavily on performance assessments that are criterion-referenced. These performance measures should compare student performance against

a clearly defined curriculum. But since we live in a complex world where mobility and diversity are the reality, administrators also need norm-referenced comparisons of their students' performance.

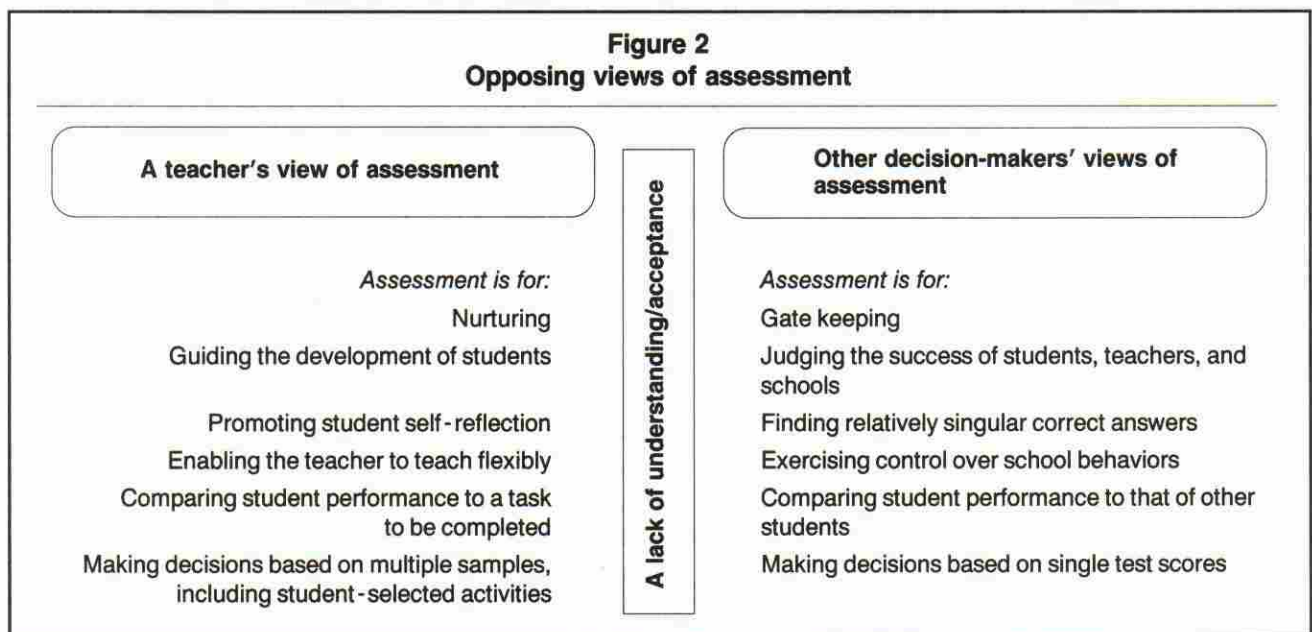
Parents. While parents share the public's interests, they have a vested interest in their own individual children. In order to monitor their children's progress and to be active in their education, parents want criterion-referenced reports; additionally parents are also typically interested in how their children perform on normed tests in comparison to children from across the United States.

Teachers. A teacher's primary concern is helping students learn. While teachers are necessarily aware of normed assessment's

Figure 1
Assessment audiences

Audiences	The information is needed to:	The information is related to:	Type of information	When information is needed:
General public (and the press)	Judge if schools are accountable and effective	Groups of students	Related to broad goals; norm- & criterion-referenced	Annually
School administrators/ staff	Judge effectiveness of curriculum, materials, teachers	Groups of students & individuals	Related to broad goals; criterion- & norm-referenced	Annually or by term/semester
Parents	Monitor progress of child, effectiveness of school	Individual student	Usually related to broader goals; both criterion- & norm-referenced	Periodically; 5 or 6 times a year
Teachers	Plan instruction, strategies, activities	Individual student; small groups	Related to specific goals; primarily criterion-referenced	Daily, or as often as possible
Students	Identify strengths, areas to emphasize	Individual (self)	Related to specific goals; criterion-referenced	Daily, or as often as possible

Figure 2
Opposing views of assessment



comparative reports as a kind of bottom-line accountability, they are primarily interested in the kind of information that will support the daily instructional decisions they need to make. This kind of information has been generated by criterion-referenced tests and by other types of assessment that can be utilized more effectively in the classroom as a part of instruction.

Students. Students need to become good self-assessors if they are to improve their literacy skills. They need to select, review, and think about the reading and writing they are doing. They need to be able to revise their own writing and to revise their comprehension as they read. If students understand their own needs, they will improve. Students should, in fact, be the primary assessors of their own literacy development.

The wall between understanding

It is important for each of these audiences to recognize, understand, and respect the needs of the others if we are to pull the assessment puzzle together. Audience needs cluster around those of teachers and students on the one hand and those of other decision-makers on the other.

The assessment needs of these two general groups tend to be dramatically different and even contradictory, and if the users of assessment do not recognize one another's needs, it is because these distinctions create a

kind of wall depicted in Figure 2. It is essential that we breach that wall if we are to get our assessment act together!

Some tests attempt to do it all

No single assessment can serve all the audiences in need of educational performance information. Yet developments in standardized tests have attempted to do so. The tests have added criterion-referenced interpretations, special interpretations for teachers, special reports for parents, individual score reports, and instructional support materials of various kinds. These developments have made the tests longer, more expensive, more time-consuming, and more confusing. Consequently, teachers are expected to justify these investments by making more instructional use of the test results.

At the same time, the increased investment in assessment time and money has tended to give these tests even more importance in determining school accountability and in making high-stakes educational decisions. Specifically, four potential problems have arisen.

Teaching to the test. As accountability became more and more of a concern, teachers have felt pressured to place emphasis on what the standardized tests covered, regardless of what the school curriculum called for. Over time, reading curricula have begun to reflect the skill breakdown of many tests, and reading

textbooks have tended to emphasize the skills tests cover as well.

Contaminating the evidence. Standardized reading tests used to mean something. They were genuine indications that a student who performed adequately on them could read. This was so because they *sampled* reading behavior. But now that indication is contaminated. If teachers are deliberately stressing the sub-behaviors that they know are on the tests, the assessments are no longer sampling reading behavior—they are, in effect, covering a very limited definition of it. A good score on a standardized reading test no longer indicates that the student can read in general. It means only that the student can do those limited things the test covers.

Crunching objectives. Attempts to make reading assessment tests more encompassing have tended to make them much longer. Even so, tests are forced to cover the numerous subskills they contain with only a few items each. “What does it mean,” a teacher may legitimately ask, “if a student misses one of three items that report on comprehending cause-and-effect?”

The potential for a mismatch. Teachers have long noted that nationally normed tests do not reflect particular emphases in their classrooms. How can a standardized reading test, they have correctly argued, tell them much about a particular curriculum they are following? What can it tell the public about how well the teacher has done using the curriculum?

The more a teacher adheres to instruction related directly to the needs, interests, and backgrounds of his or her particular students, the less assured is the match of that instruction to standardized test content—and the less likely the test’s scores will serve that instruction.

Good reading theory recommends authentic performance assessment

Most published tests have not adequately responded to emerging reading theory, which explains reading comprehension as a meaning-constructing process. Any subskills factored out of the process are not discrete; if they actually exist as behaviors, they appear to operate in such an intricate fashion that it is difficult if not impossible to isolate them.

Authentic assessment. Relatively wide acceptance of a constructivist, context-spe-

cific definition of reading has promoted a careful analysis of current reading and language arts test content and format to see how authentic the testing experience is. This analysis has led to the conclusion that the reading required on most tests is not much like the reading behavior that our new understanding describes. How valid is the content of a reading test in terms of reader purpose, interests, and background, which we now believe are primary influences on reading behavior?

Performance assessment. Attention to authenticity has accompanied and helped generate the development and use of performance assessment. A student’s language behaviors need to be assessed, it is contended, as they are used in real-life situations. Students don’t comprehend something read, for example, as a multiple-choice response, and marking those answers has nothing to do with the way reading is actually used, except in taking tests. Reading performance assessment must look at the reading act in process or judge comprehension of a text as it is applied in some realistic way.

Observation. Observation is one way to do this and can lead teachers to meaningful insights about the progress and needs of individual students. Yet teachers need to be trained in regard to what they can look for and what those signs suggest. They need to develop useful ways to make discrete notes about observations and to synthesize what they find. Observation generates many details in relatively random order, and they seldom become clearly useful until they are gathered into patterns that can direct instruction.

Portfolios. Another highly valuable form of performance assessment is the portfolio. For these collections, students and teachers select numerous samples from drafts and final versions of various kinds of a student’s writing. The idea is to demonstrate the student’s progress and development in the combined process of reading, thinking, and writing. Thus many of the samples in the portfolio are responses to reading. The portfolio is reviewed and discussed regularly by the teacher and student, who may arrange it for others to examine.

Integrated assessment. Assessments in which thinking, reading, and writing are integrated have been developed in recent years. Such assessments have been developed by

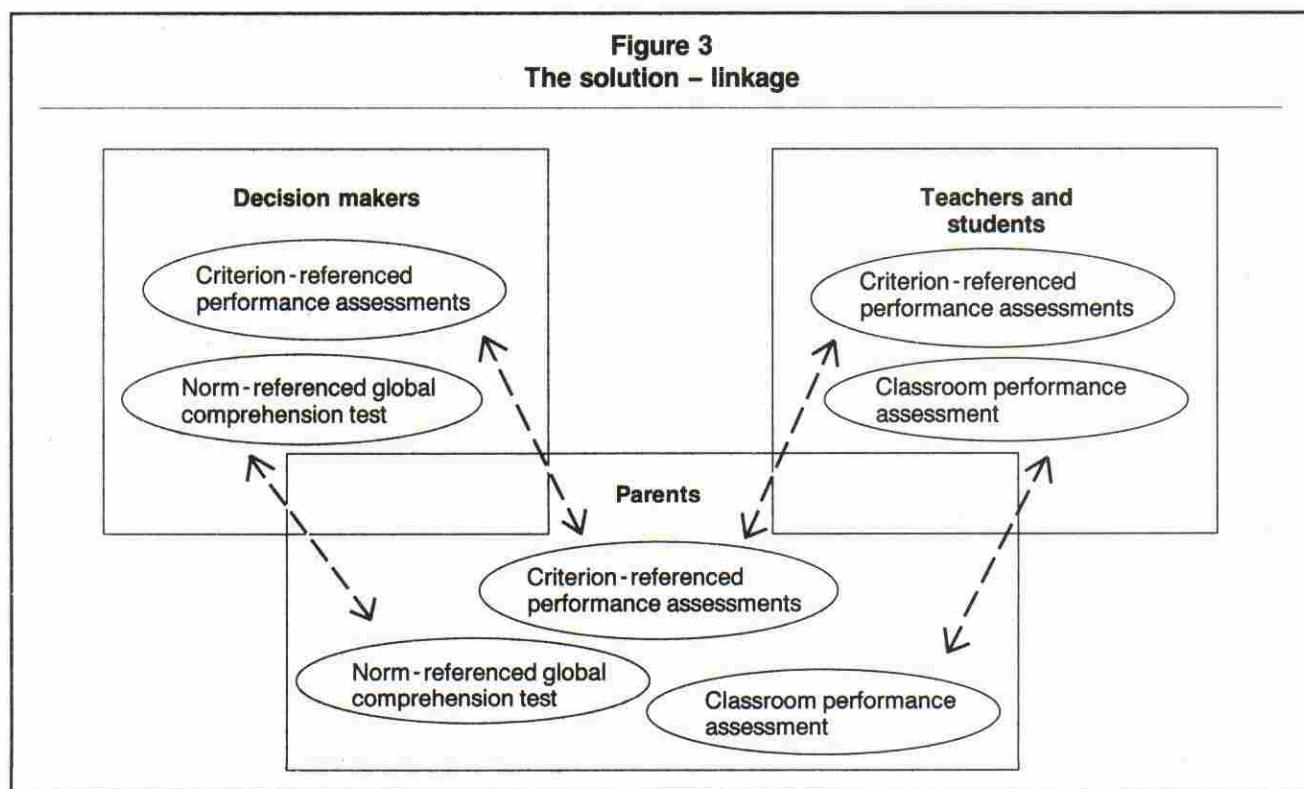
classroom teachers, school districts, and publishers in an attempt to integrate reading and writing and to assess reading and writing with more realistic activities. These vary widely, but for most of them the student is given a writing task related to a text that is supplied. The task has been deemed to be authentic because it is typical of something the student might do in real life, including the kinds of activities often used for learning in the classroom. It is designed to emphasize the use of information in the reading selection in a realistic and interesting writing task.

For example, one such test asks students to read a nonfiction article that categorically discusses and describes how insect-eating plants lure, capture, and digest their victims. The task is to write a fictional piece telling what a mother bug might say to her children in cautioning them about these plants. Teachers use what the students write to assess students' understanding of the text. They rate other integrated behaviors as well, such as the students' organization and application of the text's content to the task and factors related to writing.

Such reading/writing assessments encourage students to develop a variety of responses based on their interpretation of the reading selection, their background knowledge, and the direction they choose to take in constructing a realistic response. These kinds of performance assessments provide teachers with valuable insights regarding a student's ability to read, write, and construct a meaningful response to an interesting task. Prewriting notes, first drafts, and teacher observation notes all make the assessment a valuable source of information.

In addition, the final drafts can be scored to serve as information that can help determine accountability. The responses can be scored following a "rubric," a list of criteria that describes several levels of performance in each of the categories to be analyzed. Samples of actual student papers ("anchors") that represent each score level described by the rubrics can also be used in scoring. Thus these tests are criterion-referenced. Yet the guides to scoring are somewhat equivalent to normed scores in the sense that the anchor papers were taken from many gathered in field-testing and

Figure 3
The solution - linkage



were judged to be typical of the range of responses described in the rubric.

A combined solution to the assessment puzzle

None of the preceding types of assessment should be argued to be the single solution to the testing puzzle. Figure 3 depicts how performance assessments can provide direct linkage among the main users of assessment and how the three major types of assessment are linked. The chart is a plan for pulling the pieces of the assessment puzzle together into a solution that can inform all the decision makers involved in a student's development into an effective reader and language user.

Solving the puzzle will require cooperation

Pulling the assessment puzzle together will require tolerance and compromise on the part of many critics of particular types of assessment. The process would be facilitated if:

- Critics of the schools would become aware that assessment must serve more than school accountability. Ideally, critics will inform their concerns with a better understanding of what schools are trying to accomplish.
- Decision makers would understand that assessment is more than numbers on a test paper. They would begin to understand and use the kinds of assessments that are based on real classroom activities and that represent the types of activities in which students who are effective readers and writers should become proficient.
- The most idealistic of the critics of assessment would become more realistic and flexible, tempering their insistence on authentic performance assessment. It seems fruitless, in particular, for some critics to insist that all assessment revolve around observation of activities that are apt not to involve all children and that reveal language use in highly varying degrees.
- Producers of assessments would acknowledge that no one assessment is going to suffice as a school's examination of reading. This would mean that they would no longer promote any of

their products as such a test. It would also mean that future revisions of standardized reading tests would undo much of the complexity they now contain.

None of this is to suggest that critical analysis of reading assessment should stop, nor should attempts to improve tests in response to criticism cease. Efforts to develop and institute the new accountability assessments in Illinois (Pearson & Valencia, 1987), where the assessment allows for multiple correct responses within each multiple-choice item, and in Michigan (Michigan State Board of Education, 1987), where the assessment relies on longer passages followed by more numerous items, have been interesting, if not conclusive, efforts to contribute to a solution to the assessment puzzle. So have attempts to construct items that will reveal students' awareness of how they are processing texts. Although longer reading test passages, different question formats, etc. will not solve the assessment puzzle, they can certainly shape the parts we pull together for a better fit.

Pulling the assessment puzzle together will require tolerance and compromise on the part of many critics of particular types of assessment.

Norm-referenced tests need to change

To solve the assessment puzzle, it will be necessary for teachers and other educators to admit that norm-referenced test results can be of some value to the public and other decision makers, including parents. But these standardized tests should not be of the form that has evolved in response to criticism.

Test authors and publishers should begin to plan assessment *programs* that address the multiple audiences. Teachers and schools will need assistance in developing portfolios, planning performance assessments, and integrating assessment information. What is not

needed are large single test batteries that promise to meet all of a school's assessment needs from classroom diagnosis to accountability. That attempt, especially linking accountability assessment and instructional assessment, has led to a narrowing of the curriculum.

For the large-scale assessments, this suggests the elimination of the designation of items by subskills and reporting on those subbehaviors as if they truly are separable and distinct. More publisher efforts should go into the development of a variety of creative and useful curriculum assessments in which students have to actually perform the behaviors the school is attempting to teach.

What large-scale assessment can and should do is to report a global comprehension score, with no special subtests on traditional focuses like word recognition and vocabulary. Without the time-consuming battery of accompanying tests, reading tests can be shorter while using longer passages of a variety of types. These passages must evoke different purposes for reading that reflect the real reasons students read in and out of school. Thus, the reading test will be more authentic.

Without the burden of reporting on a host of specific reading and thinking subskills, test makers can write items that truly reflect the balance of a passage, the students' probable purpose for reading such a text, and the aspects of the writing that make the text one of quality and worth the students' time.

It should also be remembered that the long-standing primary purpose of large-scale testing has been to provide a general assessment as to how groups of students are progressing in a school district. Such information, if it does not become the focus of instruction, can be one piece of information used to contribute to a broad base of information for planning, supporting, and evaluating school- and system-wide curricula and instruction.

This approach strongly suggests that *matrix sampling* be used for large-scale assessment, thus eliminating the need to administer to all students the same test items or tasks. Testing time can be considerably shorter if carefully selected samples of students take different parts of a test instead of the whole thing. Good sampling should yield results

similar to those obtained when all students take the entire test. Nothing is lost in reporting, since individual scores are of little concern. In addition, matrix sampling provides a general indication of the progress of groups of students, not a blueprint for instruction of individual students.

Performance assessments can provide the key linkage

Figure 3 illustrates the linkages across three general audience types that will be essential to solving the assessment puzzle. Norm-referenced information provides a link between parents and decision makers other than teachers. However, the key linkage across all three general audiences is criterion-referenced performance assessments. Various approaches to performance assessment are being developed and tried out in school district assessment programs. Such assessments can be designed by teachers themselves. In fact, this has been done in several local school districts around the United States by teachers cooperating and interacting in order to meet their assessment needs. The same procedures are being tried at the state level in Maryland, Arizona, California, and Utah, and other states are sure to move in this direction.

The teachers who have been most successful in using this approach have had the support of administrators who could see over the assessment wall. Their support generated public interest and support. In some school systems, published or teacher-created integrated language performance assessment has already become a primary source of information for judging school accountability.

While teachers can create integrated language performance activities on a classroom basis, using them for accountability will require carefully developed or prepared programs that have been made congruent system-wide. This was done in River Forest, Illinois, where teachers developed their own rubrics, anchor papers, and inservice training. This kind of structuring will be necessary if the public, the press, and administrators are to be expected to value these tests as the key indicators of accountability and sources of directions for key decisions, such as curriculum development.

At the same time, of course, these tests can reflect authentic student performance.

Not only are they very closely related to instructional activities and thus of high utility to teachers, they are actually instructional activities in and of themselves so the class time they require is doubly well invested.

The portfolio is the flagship of performance assessment

Most developers of integrated language assessment programs highly recommend putting the student products into portfolios, a direct acknowledgment that the roots of language performance assessment lie in a portfolio approach to assessment and instruction. So integral is the portfolio performance assessment in good classrooms today that it is vital to note the qualities that make the portfolio approach a successful one.

A successful portfolio approach to assessment must revolve around regular and frequent attention to the portfolio by the student and the teacher. It does minimal good just to store a student's papers in a big folder and let them gather dust for lengthy periods of time. Papers must be added frequently; others can be weeded out in an ongoing rearrangement and selection process; most importantly, the whole process should involve frequent self-analysis by the student and regular conversations between the teacher and the student.

Too many teachers who contend that they are using portfolios do not do these things. Here are a few requirements if portfolios are to provide good assessment:

- The portfolio *belongs* to the student. It is his or her work and property, not some classroom requirement. Students should have choice about what goes in, and they should be encouraged to decorate and personalize their portfolios in unique ways.
- Portfolios are not primarily a display, although students may help arrange them for their parents and administrators to see. They are a shifting, growing repository of developing processes and ideas—a rather personal melting pot that the student uses to reflect on his or her own literacy development and to discuss interesting reading and writing activities with the teacher.
- The teacher's role in portfolio development is that of a consultant who helps

convince the student that the work should show a variety of materials reflecting the reading-writing-thinking process as well as examples of responses to common classroom tasks and the student's favorite creations.

- The portfolio should contain numerous and varied pieces written and revised in response to reading. Reading logs reporting ongoing responses to books and articles make valuable contributions to portfolios.
- Portfolios should be reflective collections, revealing genuinely individual and personal responses to classroom activities and to ideas.
- At an absolute minimum, there should be four one-on-one, teacher/student discussions and analyses each semester of a student's developing portfolio. These sessions should not be short and perfunctory. If this requirement is not met, the assessment potential of the portfolio process is forfeited.
- Keeping the portfolio is an ongoing process. Its real value as an assessment tool materializes as the student can analyze his or her progress and development over time.

New emphases in assessment have common qualities

Portfolios are part of a group of classroom performance assessments, some of them quite informal, that link the assessment interests of teachers, students, and parents. Portfolios can also be highly revealing to school specialists and administrators who, with the students' permission, take the time to examine them. All of these emerging strategies are both authentic and involve performance assessment. They are:

- Highly individualized, even though they may take place during activities that involve groups of students.
- A part of classroom activities and instruction designed to match an individual student's interests and needs and to use a student's strengths to develop more incisive and creative use of language.
- Activities that integrate several language behaviors.
- Chances to use critical thinking and to

express unique and emerging reactions and responses to ideas encountered in text.

- Models that encourage and develop self-assessment by the student, making him or her aware of the language-related strengths that are developing.

How school districts can begin to solve the assessment puzzle

Too often school district testing programs are nothing more than test-and-file procedures. The tests are administered; when the scores are available, they are reported in some way; and teachers are admonished to peruse and use the test results. Yet many educators across the U.S. already embrace the suggestions made here for solving the assessment puzzle. Administrators are aware that testing programs can and do divide educators.

The assessment puzzle can be solved. The solution, however, is not as simple as identifying a nonexistent test that will do the whole job nor as arbitrary as eliminating most reading assessment.

Superintendents do not want to abandon their accountability responsibilities, yet they want to support effective ongoing classroom assessment that provides teachers with information that is congruent with current knowledge about reading/writing processes. Teachers want to be more involved in developing an assessment program that serves and matches their instructional needs. They all sense that what is needed is an integrated system that is effective in fostering better teaching and learning.

Many of these school districts need help with developing an assessment program that links audiences instead of dividing them—one that supplies broad-based accountability information yet is customized to the particular

system, its teachers, and its students. One way for school districts to begin is to discuss the pieces of the assessment puzzle in their system. Representatives of all the audiences with assessment needs should take part. As this process develops, the discussions need to be recorded in some way and synthesized. Out of all this can come other brainstorming sessions and ultimately inservice workshops to help all teachers understand how a broad-based assessment program can be pulled together. Equally important, many teachers will welcome inservice training on using different types of informal assessments.

These kinds of workshops can be started within school districts right away. For instance, teachers who are exceptionally good observers or use the portfolio approach with great success are almost always easily identified. They could be enlisted and supported by administrators to run workshops that can be conducted while the discussions about broader reading assessment are helping representative groups define the assessment problems and their district's needs.

The assessment puzzle can be solved. The solution, however, is not as simple as identifying a nonexistent test that will do the whole job nor as arbitrary as eliminating most reading assessment. Rather it takes a vision that focuses on what real literacy means and the awareness that various groups have a stake in helping students to develop as literate citizens. Such a vision must not use assessment to isolate. It must respect the complex nature of literacy, it must serve students and help them to become reflective self-assessors, and it must create links that bring instruction and assessment together.

References

- Brown, R. (1986). Evaluation and learning. In A.R. Petrosky & D. Bartholomae (Eds.), *The teaching of writing: Eighty-fifth yearbook of the National Society for the Study of Education* (pp. 114-130). Chicago, IL: University of Chicago Press.
- Burstall, C. (1986). Innovative forms of assessment: A United Kingdom perspective. *Educational Measurement: Issues and Practice*, 5, 17-22.
- Bush, G. (1991). *America 2000: An education strategy*. Washington, DC: U.S. Department of Education.
- Cambourne, B., & Turbill, J. (1990). Assessment in whole language classrooms: Theory into practice. *The Elementary School Journal*, 90, 337-349.
- Farr, R. (1970). *Reading: What can be measured?* Newark, DE: International Reading Association.
- Farr, R., & Carey, R. (1986). *Reading: What can be measured?* (2nd ed.). Newark, DE: International Reading Association.

- Farr, R., & Fay, L. (1982). Reading trend data in the United States: A mandate for caveats and caution. In G. Austin & H. Garber (Eds.), *The rise and fall of national test scores* (pp. 83-141). New York: The Academic Press.
- Farr, R., Fay, L., Myers, R., & Ginsberg, M. (1987). *Reading achievement in the United States: 1944-45, 1976, and 1986*. Bloomington, IN: Indiana University.
- Finn, C.E., Jr. (1992, January 12). Turn on the lights. *The New York Times*, Sect. 4, p. 19.
- Goodman, Y. (1991). Evaluating language growth: Informal methods of evaluation. In J. Flood, J. Jensen, D. Lapp, & J. Squire (Eds.), *Handbook of research on teaching the English language arts* (pp. 502-509). New York: Macmillan.
- Jongsma, K. (1989). Portfolio assessment. *The Reading Teacher*, 43, 264-265.
- Madaus, G.F. (1985). Public policy and the testing profession: You've never had it so good? *Educational Measurement: Issues and Practice*, 4, 5-11.
- McClellan, M.C. (1988). Testing and reform. *Phi Delta Kappan*, 69, 766-771.
- Michigan State Board of Education. (1987). *Blueprint for the new MEAP reading test*. Lansing, MI: Author.
- Mullis, V.S., Owen, E.H., & Phillips, G.W. (1990). *Accelerating academic achievement: A summary of the findings from 20 years of NAEP*. Princeton, NJ: Educational Testing Service.
- National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Department of Education.
- Pearson, P.D., & Valencia, S. (1987). *The Illinois State Board of Education census assessment in reading: An historical reflection*. Springfield, IL: Illinois State Department of Education.
- Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Priestley, M. (1982). *Performance assessment in education and training: Alternate techniques*. Englewood Cliffs, NJ: Educational Technology Publications.
- Purves, A., & Niles, O. (1984). The challenge to education to produce literate citizens. In A. Purves & O. Niles (Eds.), *Becoming readers in a complex society: Eighty-third yearbook of the National Society for the Study of Education* (pp. 1-15). Chicago, IL: University of Chicago Press.
- Resnick, D. (1982). History of educational testing. In A.K. Wigdor & W.R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies*, Part 2 (pp. 173-194). Washington, DC: National Academy Press.
- Salganik, L.H. (1985). Why testing reforms are so popular and how they are changing education. *Phi Delta Kappan*, 66, 628-634.
- Salmon-Cox, L. (1981). Teachers and tests: What's really happening? *Phi Delta Kappan*, 62, 631-634.
- Saretsky, G. (1973). The strangely significant case of Peter Doe. *Phi Delta Kappan*, 54, 589-592.
- Stedman, L.C., & Kaestle, C.F. (1987). Literacy and reading performance in the United States from 1880 to the present. *Reading Research Quarterly*, 22, 8-46.
- Stiggins, R.J., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, 5, 5-17.
- Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. *The Reading Teacher*, 43, 338-339.
- Valencia, S., & Pearson, P. (1987). Reading assessment: Time for a change. *The Reading Teacher*, 40, 726-732.
- Wolf, D.P. (1989). Portfolio assessment: Sampling student work. *Educational Leadership*, 46, 35-39.

UNICEF: A new Halloween tradition

Halloween—the very mention of it shifts a child's imagination into high gear. U.S. teachers can tap into that Halloween energy by inviting children to trick-or-treat for UNICEF, the United Nations Children's Fund.

The practice began in 1950 when youngsters in Philadelphia volunteered to go door to door and raise money for needy children around the world. They collected \$17 and launched a new Halloween tradition. By presidential proclamation, October 31 has been recognized as National UNICEF Day in the United States.

UNICEF collections have become a major source of support for children in need. Last year alone, American children raised over \$2 million to provide safe drinking water, essential vaccines, health care, education, and emergency relief for the world's poorest children in 128 developing countries.

Teachers who encourage students to participate in this year's collection will help children around the world while leading their students to a more global point of view. "National UNICEF Day is a great way to introduce students to other cultures," said Lawrence E. Bruce, Jr., president of the U.S. Committee for UNICEF. "Students cross the multicultural boundaries and also learn that they can actually help save lives."

This year the U.S. Committee for UNICEF has developed a teacher's guide, *Global Awareness in the Classroom: An Introduction to Mexico*, as part of a classroom packet. Mexico was chosen in recognition of the 500th anniversary of Columbus's historic voyage. *Global Awareness* offers a brief introduction to the history and modern culture of Mexico and outlines activities for young learners.

To participate in the 1992 National UNICEF Day campaign and to obtain a curriculum guide and trick-or-treat-for-UNICEF materials (including the traditional orange collection cartons and Halloween safety tips), write the U.S. Committee for UNICEF, P.O. Box 182248, Chattanooga, TN 37422-7248, USA, or telephone 1-800-252-KIDS.

Copyright of Reading Teacher is the property of International Reading Association 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.