University of Nebraska - Lincoln

# DigitalCommons@University of Nebraska - Lincoln

November 1998

# Two Steps Forward, Three Steps Back: The Stormy History of Reading Comprehension Assessment

Loukia K. Sarroub
*University of Nebraska-Lincoln*, lsarroub@unl.edu

P. David Pearson
*Michigan State University*

# Two Steps Forward, Three Steps Back: The Stormy History of Reading Comprehension Assessment

## LOUKIA SARROUB and P. DAVID PEARSON

*The only freedom that is of enduring importance is freedom of intelligence, that is to say, freedom of observation and of judgment exercised in behalf of purposes that are intrinsically worthwhile.*

———John Dewey

After closely examining the recent history of reading comprehension assessment in the United States, we have concluded that although both the forms of assessment and the key players in the assessment process have changed in significant ways, the functions of assessment have remained relatively constant. In terms of function, we have always used, and continue to use, assessment tools to evaluate programs, to hold particular groups accountable for some specified set of outcomes (though it may seem that that is *all* we do these days), to inform instruction, either for individuals or whole classes, and finally, to determine who gains access to particular programs or privileges (the gatekeeping function). However, very different test formats, or at least a very different mix of formats, are used today than were used twenty-five years ago. We contend that changes in our fundamental views of the reading process have paved the way for these new formats. We argue that changing and sometimes conflicting policy contexts (what legislators and other policymakers want from assessments) have been responsible for shifting an emphasis from some functions (e.g., instructional decision making) to others (e.g., ac-

*Loukia Sarroub is a Ph.D. candidate in the Department of Teacher Education at Michigan State University, with concentrations in both educational policy and social analysis and literacy education. P. David Pearson holds the John A. Hannah Distinguished Professorship of Education in the College of Education at Michigan State University, where he is a member of the Department of Teacher Education and the Department of Counseling, Educational Psychology, and Special Education.*

countability) and have changed who it is that decides who shall take what tests and for what purposes.

We also attempt to document another thesis, one that is more interpretive than descriptive: Progress, if one can even characterize the history of reading comprehension assessment as moving in a particular reform-minded direction, is best characterized as "two steps forward, three steps back." Usually, a forward step is an advance in assessment practice driven by an advance in reading theory, or possibly psychometric theory. Usually, the backward step is a retreat in assessment practice driven by some political or practical constraint. As we discuss later, the most notable retreat in the last quarter-century has been in the area of accountability. In the name of holding schools and teachers responsible for student performance, education officials have created such a high-stakes environment that people end up "teaching to the test" in a way that narrows rather than expands curricular opportunities. A second "step back" has been the retreat in the use of portfolios and performance assessments; they are considered either too personal (a political motive) or too time-consuming for the quality of information obtained (a practical motive).

We make these two points by examining the historical course of reading comprehension assessment practices over the last quarter-century. To understand the current mix of comprehension assessment practices, we believe that it is necessary to begin with a characterization of the assessment practices that were dominant in the 1970s and then to work our way to the present, trying to understand each new assessment twist in light of changing views of reading processes, practices, and policies.

### Reading Assessment in the Early to Middle 1970s

Before we examine reading assessment in the 1970s, we are obligated to present a short history of reading comprehension assessment prior to that decade. The history is short because we have not long been assessing reading compre-

97

hension. In fact the first systematic attempts to index read-ing ability by measuring comprehension date back to World War I. Thorndike, in his classic 1918 piece "Reading as Reasoning," offered us our first professional glimpse inside the mind of the reader; Thorndike tried to characterize what *must* go on in the mind to produce the sorts of answers readers come up with in response to questions about what they have read. The quest to get as close as possible to the "phenomenological act of comprehension" as it occurs has always driven researchers to discover new and better in-dices of reading comprehension. Hard as they try, however, they have always had to settle for indirect indices of the actual process: (*a*) a short answer (which arose in the teens and twenties), (*b*) bubbles filled in (which began in the thir-ties), (*c*) the essay (championed after World War II), and (*d*) an oral response in a discussion (always there but only used as an assessment tool in recent portfolio implementations). Each of these indices represents the residue of the compre-hension process, however, rather than the process itself. Furthermore, they all interpose some other cognitive and/or motor task (marking, writing, speaking, or reflecting) be-tween the act of comprehension and the evidence of its oc-currence. Even today, comprehension scholars bemoan this shortcoming in their work.

In the 1970s, there were three important facets of read-ing comprehension assessment: (*a*) standardized, multiple-choice tests, (*b*) criterion-referenced assessments of specif-ic skills, and (*c*) informal classroom assessments of comprehension. Both the multiple-choice tests and the in-formal tests had, by 1975, a long history in our schools, but criterion-referenced assessments were relatively new.

It is difficult to pinpoint the underlying theoretical ratio-nale for standardized, multiple-choice questions of reading comprehension. By 1975, they had been around for so long that few questioned their place in our educational system. It is probably fair to say that they were conceptually shaped within the prevailing behaviorist psychology that dominat-ed education from the 1930s through the 1960s. The pur-pose of the tests was to tell how well a student did in rela-tion to a national sample of scores. Students took tests, teachers sent them off for scantron scoring, and several months later teachers, students, and parents received some sort of score report indicating either a grade norm score (Johnny scored at 6.3) or a national percentile (Johnny scored at the 35th percentile). Figure 1 shows a portion of a standardized, multiple-choice test for fourth graders in the 1970s.

Criterion-referenced assessments (i.e., students were judged to have mastered the phenomenon being tested if and only if they achieved a certain cut score on the test) were the products of a new and exciting emphasis on mas-tery learning prompted by conceptual advances in under-standing the relationship between learning and assessment put forward by scholars such as Benjamin Bloom (1968) and Robert Gagné (1965). The idea was that if we could just be more precise about the essential elements involved

---

### FIGURE 1
### Example of a Standardized, Multiple-Choice Test for Fourth Graders in the 1970s

One day last summer Mother took my cousin Tom and me to the zoo. The keeper told us about the ani-mals. The one we liked the best was the giraffe, the tallest animal in the zoo. He was 13 feet, 6 inches high from the ground to the top of his head. His front legs were about 5 feet long, and his back legs were about 4 feet long. The door to his shed was 10 feet high. His feed box was 6 feet above the floor.

12. Who is the writer of this story?
    (1) Tom
    (2) Tom's cousin
    (3) Tom's mother
    (4) The zoo keeper
13. What does the "one" stand for in the third line?
    (1) Animal
    (2) Giraffe
    (3) Keeper
    (4) Story
14. How did the children learn about the giraffe?
    (1) They read a story about a giraffe.
    (2) The studied giraffes in science class.
    (3) The keeper let them measure the giraffe.
    (4) The keeper told them about the giraffe.
15. The giraffe's front legs are how much longer than his back legs?
    (1) 9 feet
    (2) 5 feet
    (3) 4 feet
    (4) 1 foot
16. Why is the feed box high above the floor?
    (1) So other animals cannot steal the food
    (2) To help keep the floor clean
    (3) So the giraffe can stand while eating
    (4) To make it easier to fill from the top
17. How tall was the giraffe?
    (1) 6 feet
    (2) 10 feet
    (3) 13 feet
    (4) More than 13 feet

---

*Source:* Adapted from Pearson, P. D., and D. Johnson. 1978. *Teaching Reading Comprehension.* New York: Holt, Rinehart, and Winston.

---

in learning any particular domain or process, we could bring most, if not all, students to higher levels of achieve-ment. For Bloom, the question of domain was left open to definition by whoever was designing the curriculum. The precision could be achieved, according to the champions of mastery learning (see Otto and Chester 1976), by decom-posing the domain or process into essential elements and then teaching (and testing) each of the elements to mastery. For comprehension assessment, the consequences were dramatic. Even with standardized, multiple-choice assess-ments, there had been some sense that important aspects of a passage ought to be queried. With the new criterion-ref-erenced assessments of reading comprehension, the num-ber of comprehension subskills increased dramatically, as

did the number of specific skill tests for each of these. Clearly, in these assessments the emphasis is on the skill rather than the passage (for example, the exercise in figure 2 tests a child's ability to recognize sequential order).

Criterion-referenced assessments were popular throughout our schools and curricula in the 1970s and 1980s, but nowhere was their influence more dramatically felt than in basal reading materials. Starting in the early 1970s, basal programs developed criterion-referenced tests for every unit (a grouping of six to eight stories plus associated activities) and every book in the series. These tests were often used to determine who was or was not ready to go on to the next unit. In the true mastery programs (e.g., Wisconsin Design for Reading Skill Development [Otto 1977]; Chicago Mastery Learning [Board of Education 1981]) and in some basal programs, students who failed a particular subtest were required to practice skill sheets that looked remarkably like the mastery tests until they could achieve mastery (which was usually and arbitrarily defined as 80 percent correct).

Informal assessments of reading comprehension also have a long history, going back as far as the ubiquitous informal reading inventory (Kilgallon 1942, cited in Betts 1946). After students have read a passage (sometimes orally and sometimes silently), they are asked either to retell the passage (e.g., Durrell 1955) or to answer four to six comprehension questions about it. Typically, standards (e.g., percentage of ideas recalled or percentage of questions answered correctly) are set for independent, instructional, and frustration levels, just as they are for oral reading accuracy. Another type of informal reading inventory, reading miscue inventories (Goodman and Burke 1970), although it called for students to read aloud, shifted the focus from determining levels of reading to explicating a reader's comprehension and decoding strategies. Figure 3 is an example of comprehension questions that may follow the reading of a short story.

Another important legacy of this period, which had nothing to do with assessment content or format, was the reappearance and gradual acceleration of an accountability mentality. Accountability was not new in the 1970s. The initial "foot in the door" for accountability came with the reauthorization of Title I in 1968. In that reauthorization, states and districts essentially made a deal in which they traded accountability (a promise to strive for particular growth targets on standardized tests) for dollars that they could use to help struggling readers. The second foot came through the door with the development and spread of state assessment systems in the early 1970s. The public reporting of state assessment data on a district-by-district or school-by-school basis brought us the notion of "high-stakes assessment," assessments that were so important to schools and teachers that they spent inordinate amounts of time and energy getting students ready to take those tests, just so their schools would look good—or at least so they would not look too bad (Pearson and Dunning 1985).

---

**FIGURE 2**
**A Passage Designed to Test a Student's Ability to Recognize Sequential Order**

The children wanted to make a book for their teacher. One girl brought a camera to school. She took a picture of each person in the class. Then they wrote their names under the pictures. One boy tied all the pages together. Then the children gave the book to their teacher.

1. What happened first?
   a. The children wrote their names.
   b. Someone brought a camera to school.
   c. The children gave a book to their teacher.
2. What happened after the children wrote their names?
   a. A boy put the pages together.
   b. The children taped their pictures.
   c. A girl took pictures of each person.
3. What happened last?
   a. The children wrote their names under the pictures.
   b. A girl took pictures of everyone.
   c. The children gave the book to their teacher.

*Source:* Adapted from *The Ginn Reading Program.* Lexington, Mass., 1982.

---

**FIGURE 3**
**Examples of Comprehension Questions**

Harriet and Uncle Bill got into the little blue airplane. Soon the plane was ready to go.
"Let's go," said Uncle Bill.
Then they took off. Up, up they went. Soon they were way up in the air.
"Look down there," said Uncle Bill.
"That's Red Rock Ranch."
Harriet looked down at the ranch. The cows looked very, very little. The ranch house didn't look as big as a play house.
Uncle Bill said, "Hold on. Here we go!"
The airplane went around. Up, up, up—then over and down. Over and over it went.
"Oh, Uncle Bill," said Harriet.
"I feel funny. Please don't go over and over like that."
Uncle Bill laughed. "All right," he said.
"No more tricks. Now let's land the plane."

1. Was Uncle Bill's plane big or little?
2. What were some of the things Harriet saw from Uncle Bill's plane?
3. Why did the cows and ranch look so small?
4. Why do you think Uncle Bill told Harriet to "hold on" when he made the airplane do tricks?
5. How do you think Harriet felt when Uncle Bill made the airplane do tricks?
6. Did Harriet like her airplane ride? Why do you think so?

*Source:* Cooper, J. D. et al. 1972. *Decision Making for the Diagnostic Teacher.* New York: Holt, Rinehart, and Winston.

## The First Revolution in Assessment

By the mid-1980s, we were in the midst of the cognitive revolution in teaching reading (as it turned out, this revolution was merely a prelude to the sociocultural revolution that lay just around the corner). Figure 4 illustrates the tensions that existed between the new cognitively oriented views of the reading process and the assessment praxis at that time.

By the late 1980s, constructivist approaches to reading assessment included the need to rely on resources such as prior knowledge, environmental clues, the text itself, and the key players involved in the reading process. Even then, reading was thought to involve a reflective element. The complementary notion that strategic reading (rather than skilled reading) necessitated assessments that were consis-

tent with that goal became the prevalent argument among educators in the policy arena. Unfortunately, as much as teachers might have wanted to focus their attention on reading assessments that did not focus simply on norm or criterion-referenced skills, they still wanted to help their students become successful readers by the most public and widely accepted of standards. The stakes, as we have suggested, were high, and tests were often convenient, if not necessary, guides for instruction. This often meant that teachers taught to the test (Valencia and Pearson 1987). The inappropriate use of tests for instruction probably advanced unsuitable instructional models for reading and resulted in what Haladyna, Nolan, and Hass (1991) call score pollution—the phenomenon of an increase in test scores without a parallel increase in the underlying cognitive process that

---

**FIGURE 4**
**Contrast Between Cognitively Oriented Views of Reading and Current Assessment Practices**

| NEW VIEWS OF THE READING PROCESS TELL US THE FOLLOWING: | YET WHEN WE ASSESS READING COMPREHENSION, WE DO THE FOLLOWING: |
|---|---|
| Prior knowledge is an important determinant of reading comprehension. | Mask any relationship between prior knowledge and reading comprehension by using lots of short passages on lots of topics. |
| A complete story or text has structural and topical integrity. | Use short texts that seldom approximate the structural and topical integrity of an authentic text. |
| Inference is an essential part of the process of comprehending units as small as sentences. | Rely on literal comprehension of text items. |
| The diversity in prior knowledge among individuals as well as the varied causal relations in human experiences invite many possible inferences to fit a text or question. | Use multiple-choice items with only one correct answer, even when many of the responses might, under certain conditions, be plausible. |
| The ability to vary reading strategies to fit the text and the situation is one hallmark of an expert reader. | Seldom assess how and when students vary the strategies they use during normal reading or studying or when the going gets tough. |
| The ability to synthesize information from various parts of the text and different texts is a hallmark of an expert reader. | Rarely go beyond finding the main idea of a paragraph or passage. |
| The ability to ask good questions of text, as well as to answer them, is a hallmark of an expert reader. | Seldom ask students to create or select questions about a selection they may have just read. |
| All aspects of a reader's experience, including habits that arise from school and home, influence reading comprehension. | Rarely view information on reading habits and attitudes as being important information. |
| Reading involves the orchestration of many skills that complement one another in a variety of ways. | Use tests that fragment reading into isolated skills and report performance on each. |
| Skilled readers are fluent; their work identification is sufficiently automatic to allow most cognitive resources to be used for comprehension. | Rarely consider fluency as an index of skilled reading. |
| Learning from text involves the restructuring, application, and flexible use of knowledge in new situations. | Often ask readers to respond to the text's declarative knowledge rather than to apply it to near and far transfer tasks. |

*Source:* Adapted from Valencia, S., and P. D. Pearson. 1987. Reading Assessment: Time for a Change. *Reading Teacher* 40:731.

the test is supposed to measure. In other words, kids get better scores but are not better readers.

Test-score pollution led educators to change the field of comprehension assessment once again. By the late 1980s, reading comprehension tests included longer text passages, more-challenging questions, different question formats (such as multiple-multiple-choice, with more than one right answer [see figure 5], and open-ended questions), and response to literature formats. The reading field acknowledged that while all multiple-choice items include answers that are plausible under certain conditions, they do not necessarily invite reflection or interactive learning. Therefore, the intellectual shift toward the social nature of learning finally took a firm hold of the field and was realized in different ways. For example, the Illinois Reading Assessment (Illinois Goal Assessment Program 1991) promoted an interactive model of reading in which the construction of meaning became the locus around which reading strategies, dispositions toward literacy, text characteristics, and prior knowledge all revolved. Assessments in Illinois and Michigan (see Valencia et al. 1989) came to include longer and more "naturally occurring" text selections (i.e., the selections came directly from books or stories and were not written or rewritten by the test preparers).

The most significant advances in classroom comprehension assessment tools during this period also came from cognitive science. First, "retelling" spread as a way to assess comprehension. Driven by the 1970s advances in our knowledge about the structure of narrative and expository text (see Meyer and Rice 1984), many scholars (see Mitchell and Irwin, in preparation; Morrow 1985) developed systems for evaluating the depth and breadth of students' text understandings based on their attempts to retell or recall what they had read. Like the formal efforts just described, there was a conscious attempt to take into account reader, text, and context factors in characterizing students' retellings. Second, "think-alouds" became widely used as a

measure of comprehension. Think-alouds had become respectable research tools by virtue of the important work on self-reports of cognitive processes popularized by Ericsson and Simon (1980). In attempting to characterize the nature of expertise in complex activities, such as chess, Ericsson and Simon learned that the best way inside the heads of the best players was to engage them in thinking aloud about the what, why, and how of their thinking and actions during the activity. This led to the use of think-alouds as both an instructional practice (Baumann, Jones, and Seifert-Kessell 1993) and an assessment practice (Farr and Greene 1992; CLAS 1994).

## The Second Revolution in Reading Assessment

We are not sure whether what happened next really constitutes a second revolution or is better thought of as an extension of the first revolution, but for the sake of emphasis, we will give it an independent status. It came so fast on the heels of the first revolution that it is hard to pinpoint its beginning point. In fact, harbingers of this sociocultural revolution, emanating from sociolinguistic perspectives (see Bloome and Green 1984) and the rediscovery of Vygotsky (see Vygotsky 1986; Wertsch 1985) were present in the early to mid-1980s, even as the cognitive revolution was exercising its muscle on assessment practices. For example, cognitively motivated teaching approaches—such as reciprocal teaching, where students took on more responsibility for their own learning by teaching each other, and process writing steps, where revision and conversation about revision delved more deeply into the social nature of reading, writing, and understanding—were used to engage students to reflect on their work and interact with others about the work.

Nowhere was the attempt to infuse social and cultural perspectives into comprehension assessment processes more transparent than in the work of the California Learning Assessment System (CLAS). The now defunct CLAS paved the way for more open assessments by emphasizing response to literature formats and the social aspects of learning. Response to literature questions articulated an open and reflective stance toward reading rather than a skills-based approach (e.g., "'If you were explaining what this essay is about to a person who had not read it, what would you say? What do you think is important or significant about it? What questions do you have about it?' and 'This is your chance to write any other observations, questions, appreciations, and criticisms of the story'" [CLAS 6–9]). Response to literature formats demanded from students that they be able to summarize, explain, justify, interpret, and provide evidence in their answers. In other words, assessment of reading comprehension reached a new high, one much more compatible with what society might expect of students in the real world. The early work of New Standards (see Pearson, Spalding, and Myers 1998) had the same goals, theoretical grounding, and format characteristics as CLAS: (a) give students a chance to show their

---

**FIGURE 5**
**A Comprehension Question with More Than One Right Answer**

What do you think that the author Patricia Edwards Clyne wanted you to learn from reading "The Army of Two"?

1. There is safety in large numbers.
2. Keep things that you may need in the future in a safe place.
3. Lighthouses and sand dunes are dangerous places to live.
4. *It takes more than strength to win a battle.
5. *Careful thinking can sometimes make things possible that seem impossible.

---

*Indicates right answers.

expertise in artifacts that reflect both the teachers' and students' social and cultural knowledge; (*b*) let the work be interesting and relevant to students' backgrounds and cultural heritages; (*c*) let the work be guided by the support of colleagues who have the students' best interests at heart; and (*d*) let the work be born of the same motives and conditions that prevail in the worlds of work and social action.

The idea that students live in multiple worlds such as home, school, and community and are expected to relate to others across contexts grew out of the sociocultural revolution in the late 1980s, although it had well-grounded historical precedents. Dewey (1938) remarked that "[a] primary responsibility of educators is that they not only be aware of the general principle of the shaping of actual experience by environing conditions, but they also recognize in the concrete what surroundings are conducive to having experiences that lead to growth" (40). Basically, Dewey did not think that learning in isolation could lead to positive outcomes. He believed that education is a social process in which individuals form a community group. In line with the idea of the social nature of learning, comprehension assessment systems such as CLAS also devised tests that focused on the interconnectedness of individual learning within the contexts of group work (figure 6 shows such an example).

In addition to changes made in test formats, the late 1980s and early 1990s saw performance assessments and portfolios become useful classroom tools. That assessments were classroom-based was important for two reasons: (1) students were to be evaluated on what they actually did in the classroom, and (2) both teachers and students could hold positions of power as they became key players in the evaluation process. In other words, performance assessments, if they were to be successful, demanded that evaluation be open, accessible, and explicit to both students and teachers (see Sarroub et al., 1997; Bisesi et al. 1998). Per-

---

**FIGURE 6**
**The Interconnectedness of Individual Learning within the Contexts of Group Work**

Now you will be working in a group. You will be preparing yourself to do some writing later. Your group will be talking about the story you read earlier. A copy of the story is provided before the group questions if you need to refer to it. Some of the activities in this section may direct you to work alone and then share with your group, and other activities may have all of you working together. It is important to take notes of your discussion because you will be able to use these notes when you do your writing.

Read the directions and do the activities described. Members of the group should take turns reading the directions. The group leader should keep the activities moving along so that you finish all activities.

You'll have fifteen minutes for these prewriting activities.

---

formance assessments epitomized the reigning sociocultural movement in education because they emphasized a personal orientation toward evidence of growth and learning rather than the more categorical skills–based approach of previous years.

However, and as with other novel approaches in comprehension assessment, portfolios came under fire as teachers and policymakers struggled with issues of external accountability and the demand that schools and school districts be able to compare students' scores (Pearson, DeStefano, and García 1998). Nonetheless, successful implementation of authentic wide-scale assessment occurred in California and Maryland (see Weiss 1994; Kapinus, Collier, and Kruglanski 1994).

## Midcourse Correction in the Mid-1990s

Although great strides were made in research on reading comprehension assessments in the 1980s, application of the new formats did not take root. It became obvious in the mid-1980s that we were building models of basic processes and models of instructional practice on a whole new theoretical infrastructure (schema theory, the centrality of the knowledge-comprehension relationship) while our assessments had not changed since the infusion of criterion-referenced tests beginning in the early 1970s. Unfortunately, the notion of "two steps forward, three steps back" was only too real, and it exposed the political nature of reform and change. Only a century before, H. G. Wells (1892) had remarked, "The examiner pipes and the teacher must dance—and the examiner sticks to the old tune. If the educational reformers really wish the dance altered they must turn their attention from the dancers to the musicians" (382). A hundred years later, we are still learning the steps to the same dance.

By the mid-1990s, the field of reading comprehension assessment had shifted its stance once again, taking a couple steps back to accommodate political pressures. Because of a conservative backlash, psychometric suspicion about new forms of assessment, utility issues (the new approaches took too long and were too expensive and difficult to score), and equity concerns (will minority students do any better or will they, in fact, do worse on those sorts of measures?), adjustments were made to assessment systems that relied mainly on longer texts and more complex and more open formats. Players from both sides of the political aisle used one or more of these criticisms to attack and eliminate assessments such as CLAS. It is interesting to note that in the wake of CLAS's demise, already validated (at great state expense and professional involvement) performance assessments in states such as Indiana and Wisconsin were shelved without seeing the light of day.

Within the political arena, educators seem to be searching for a compromise position that preserves some of the features of the reforms of the late 1980s and early 1990s while acknowledging that those efforts may have gone too far. This compromise mentality can be seen in many current

assessment initiatives: Medium-length passages and a balance of formats, about 80 percent multiple-choice and 20 percent open-ended questions, appear to be the emerging compromise. Multiple-choice questions are used to ensure high reliability, and open-ended questions are endorsed as a way to ensure that cognitive challenge is present. We even found examples of this compromise in commercial tests such as the Stanford 9 Achievement Series (1996).

> In order to provide maximum assessment opportunities of all kinds for all schools, Stanford 9 has moved beyond the ordinary boundaries of norm-referenced achievement batteries. Now you have your choice of assessment formats, whether you prefer multiple-choice items, open-ended items in which students receive partial credit for partial knowledge, writing prompts that elicit performance in one of four modes, or a combination of any or all of these (3).

Figure 7 displays a portion of an open-ended test.

Although some tests focus to a large extent on either multiple-choice or performance assessment traditions, others—for example, the NAEP, National Voluntary Reading Test, the New Standards Reference Exam, new commercial tests, and some emerging state tests—fall somewhere in between and take advantage of both traditions. The exam-

ple in figure 8 from the 1995 National Voluntary Reading Test, a test that draws from and is based on successful formats in the 1992 NAEP, illustrates this combination of traditions. The scorers of these tests use rubrics corresponding to answers that are either short constructed responses (SCRs) or extended constructed responses (ECRs).

## Current Status of Reading Comprehension Assessment

Getting to the fundamental processes of comprehension as they occur in the mind has remained a sort of "holy grail" for comprehension researchers throughout this century. Even though it eludes us, we continue in our quest to find it.

---

**FIGURE 7**
**Sample of an Open-Ended Test in the Stanford Achievement Test Booklet**

Second-grade students are asked to answer the following questions after reading "Buster the Brave" by Alexandra Wallner:

*Get the Big Picture*
1. Below are some names in the story. Write two or three things about them that you learned from the story.
   Tinker and Sam
   Buster
2. What did Buster do when Mrs. O'Malley first brought him home?
3. What was Tinker and Sam's problem? What happened to make things better?

*Take a Closer Look*
4. Think about a part of the story that you liked.
   a. What part did you like?
   b. Why did you pick that part?
5. Did Buster like Tinker and Sam when he first met them? Why do you think that?
6. This story is called "Buster the Brave." Do you think it is a good name for the story? Why do you think that?

*Be a Critic*
7. If you could change the end of the story, how would you change it?
8. Do you think this story really could have happened? What makes you think that?
9. This is where you can write anything else that you think is important about the story or anything that shows your understanding of the story.

---

**FIGURE 8**
**An Excerpt from the 1995 National Voluntary Reading Test**

1. *Initial understanding*
   Why is the little bird in the marketplace?
   [This is an SCR item.]
   What did the little bird want more than anything else?
   a. a good home
   b. to be set free*
   c. grains of rice
   d. a cheerful owner
   [This is a multiple-choice item.]

2. *Developing interpretation*
   How are the little bird and the little girl alike? Explain your answer with information from the story.
   [This is an SCR item.]
   What did the sick little girl learn from the bird?
   a. The kindness you do will be returned.*
   b. It is important to be patient in life.
   c. It can be lonely when a friend leaves.
   d. Pets are important when you are sick.
   [This is a multiple-choice item, but it can be used as an SCR.]

3. *Reader/Text connections*
   Why do you think the little bird came back each evening to the little girl's window? Use information from the story to support your ideas.
   [This could be an SCR or an ECR item.]

4. *Demonstrating a critical stance*
   What does the author do to make it very clear how sad the little bird really is?
   [This is an SCR item.]
   Would the story have been different if the young girls had not been sick? Use information from the story to explain your answer.
   [This is an ECR item.]

*Note:* Students are asked to answer questions after reading a substantial and naturally occurring passage entitled "Under the Rice Moon." SCR = short constructed response; ECR = extended constructed response.

Although tests such as the NAEP or the National Voluntary Reading Test are useful political compromises, there is no question that they signal a retreat in the field of reading and assessment research. Without question, we have lost some of the more promising and substantial elements of authentic and relevant classroom assessments, at least as policy tools that are used for wide-scale assessment. Portfolios, extended performance tasks, and assessments that revolve around the social nature of learning are still visible in writing—and even as policy instruments in wide-scale assessment. However, they are less prominent in reading, especially in the high-stakes political arena. It is somewhat ironic to note that there is substantial evidence that even the infamous CLAS tests are alive and well, in a sort of underground network, in many California classrooms and schools; apparently, teachers, administrators, and parents have found value in these assessments even as policymakers are suspicious about their merit and efficacy. Dewey (1938) once said that "The history of educational theory is marked by opposition between the idea that education is development from within and that it is formation from without" (38). That thought is very much in keeping with the events that mark the ebb and flow of reading comprehension assessments within educational spheres and in the public political world.

As we write this piece, it is clear that we are in the middle of a "three steps back" period. It is natural to ask, what the next "one step forward" will be. If the cognitive revolution helped us take the first step forward (with all the ado about authentic texts and more cognitively challenging tasks) and if the sociocultural revolution helped us take the second step forward (the inclusion of social and cultural elements into assessment), then what perspective, lying just over our theoretical horizon, will help us take the next step forward? Will it be an extension of the postmodern movement, with its emphasis on the deconstruction of text and the questioning of traditional sources of authority when deciding on the meaning of texts? Will it be a revised version of distributed intelligence (Gergen 1991; Solomon and Perkins 1998), with its de-emphasis on the individual and greater emphasis on communitarian processes for constructing meaning?

Frankly, we are not sure where this next step forward will come from. However, we are certain that there will be one. And, we are equally certain that when that step is taken, there will be political foes and nay sayers waiting to pull us back a couple of steps. Maybe the next time around, those voices of tradition will bring us back into the 1980s, which, if our account of that decade is at all accurate, would not be a bad time and place and set of assessment practices from which to build a new century's curriculum.

## REFERENCES

Baumann, J., L. Jones, and N. Seifert-Kessell. 1993. Using think alouds to enhance children's comprehension monitoring abilities. *Reading Teacher* 47:184–93.

Betts, E. A. 1946. *Foundations of reading instruction.* New York: American Book Company.

Bisesi, T., D. Brenner, M. McVee, P. D. Pearson, and L. K. Sarroub. 1998. Assessment in literature-based reading programs: Have we kept our promises? In *Literature-based instruction: Present issues, future directions,* edited by K. Au and T. Rafael. Boston: Christopher-Gordon.

Bloom, B. 1968. Learning for mastery. *Evaluation Comment, 1.*

Bloome, D., and J. Green. 1984. Directions in the sociolinguistic study of reading. In *The handbook of reading research,* edited by P. D. Pearson, R. Barr, M. L. Kamil, and P. Mosenthal, 395–421. New York: Longmans.

Board of Education, City of Chicago. 1981. *Chicago Mastery Learning Reading.* Watertown, Mass.: Mastery Education Corp.

California Learning Assessment System (CLAS). 1994. *Elementary performance assessments: Integrated English-language arts illustrative material.* Sacramento, Calif.: California Department of Education.

Cooper, J. D., L. A. Cooper, N. Roser, L. A. Harris, and C. Smith. 1972. *Decision making for the diagnostic teacher.* New York: Holt, Rinehart, and Winston.

Dewey, J. 1938. *Experience and education.* New York: Collier Books.

Durrell, D. D. 1955. *Durrell analysis of reading difficulty.* New York: Harcourt, Brace, and World.

Ericsson, K. A., and H. A. Simon. 1980. Verbal reports as data. *Psychological Review* 87:215–51.

Farr, R., and B. G. Greene. 1992. Using verbal and written think-alongs to assess metacognition in reading. Paper presented at the 15th annual conference of the Eastern Education Research Association, Hilton Head, South Carolina, March 5–7.

Gagné, R. M. 1965. *The conditions of learning.* New York: Holt, Rinehart, and Winston.

Gergen, K. 1991. The saturated self: Dilemmas of identity in contemporary life. New York: Basic Books.

Ginn and Co. 1982. *The Ginn reading program.* Lexington, Mass.: Ginn and Co.

Goodman, Y. M., and C. L. Burke. 1970. *Reading miscue inventory.* New York: Macmillan. (Subsequent editions are published by Richard C. Owen).

Haladyna, T. M., S. B. Nolan, and N. S. Hass. 1991. Raising standardized test scores and the origins of test score pollution. *Educational Researcher* 61:405–418.

Illinois Goal Assessment Program. 1991. *The Illinois reading assessment: Classroom connections.* Springfield, Ill.: Illinois State Board of Education.

Johnson, D., and P. D. Pearson. 1975. Skills management systems: A critique. *Reading Teacher* 28:757–64.

Kapinus, B., G. V. Collier, and H. Kruglanski. 1994. The Maryland school performance assessment program: A new wave of assessment. In *Authentic reading assessment: Practices and possibilities,* edited by S. Valencia, E. Hiebert, and P. Afflerbach, 255–76. Newark, Del.: International Reading Association.

Kilgallon, P. A. 1942. Study of relationships among pupil adjustments in language situations. Unpublished doctoral dissertation. Pennsylvania State University, cited in Betts, 1946.

Meyer, B. J. F., and E. Rice. 1984. The structure of text. In *The Handbook of reading research,* edited by P. D. Pearson, R. Barr, M. L. Kamil, and P. Mosenthal, 319–52. New York: Longmans.

Mitchell, J. N., and P. Irwin, P. (in preparation). The reader retelling profile. Using retellings to make instructional decisions. Manuscript submitted for publication.

Morrow, L. M. 1985. Retelling stories: A strategy for improving young children's comprehension, concept of story structure, and oral language complexity. *Elementary School Journal* 85 (5): 647–61.

NAEP Reading Consensus Project. 1992. *Reading framework for the 1992 National Assessment of Educational Progress.* Washington, D.C.: U.S. Printing Office.

National Voluntary Reading Test. 1995. *National Voluntary Reading Test.* Washington, D.C.: U.S. Printing Office.

Otto, W. 1977. The Wisconsin design: A reading program for individually guided education. In *Individually-guided elementary education: Concepts and practices,* edited by H. J. Klausmeier, R. A. Rossmiller, and M. Saily. New York: Academic Press.

Otto, W. R., and R. D. Chester. 1976. *Objective-based reading.* Reading, Mass.: Addison-Wesley.

Pearson, P. D., and D. Dunning. 1985. The impact of assessment on reading instruction. *Illinois Reading Council Journal* 13 (2): 19–29.

Pearson, P. D., and D. Johnson. 1978. *Teaching reading comprehension.* New York: Holt, Rinehart, and Winston.

Pearson, P. D., L. DeStefano, and G. E. García. 1998. Ten dilemmas of performance assessment. In *Assessing Reading 1: Theory and Practice,* edited by C. Harrison and T. Salinger, 21–49. London: Routledge.

Pearson, P. D., L. Spalding, and M. Myers. 1998. Literacy assessment in the New Standards Project. In *Assessing reading 2: Practice and Theory,* edited by M. Cole and R. Jenkins, 54–98. London: Routledge.

Sarroub, L. K., P. D. Pearson, C. Dykema, and R. Lloyd. 1997. When portfolios become part of the grading process: A case study in a junior high setting. In *NRC Yearbook, 46th Edition,* edited by K. Hinchman, D. Leu, and C. Kinzer. Chicago: National Reading Conference.

Solomon, G., and D. N. Perkins. 1998. Individual and social aspects of learning. In *Review of research in education, 24: The social organization of learning,* edited by P. D. Pearson and A. Iran-Nejad, 1–24. Washington, D. C.: American Educational Research Association.

Stanford Achievement Test. 1996. *Open-ended directions for administering.* San Antonio: Harcourt Brace.

Thorndike, E. L. 1918. Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology* 8:323–32.

Valencia, S., and P. D. Pearson, P. D. 1987. Reading assessment: Time for a change. *Reading Teacher* 40: 726–32.

Valencia, S. W., P. D. Pearson, C. W. Peters, and K. K. Wixson. 1989. Theory and practice in statewide reading assessment: Closing the gap. *Educational Leadership* 47 (7): 57–63.

Vygotsky, L. 1986. *Thought and language.* Cambridge: MIT Press.

Weiss, B. 1994. California's new English-language arts assessment. In *Authentic reading assessment: Practices and possibilities,* edited by S. Valencia, E. Hiebert, and P. Afflerbach, 197–217. Newark, Del.: International Reading Association.

Wells, H. G. 1892. On the true lever of education. *Educational Review* 4: 380–85.

Wertsch, J. 1985. *Vygotsky and the social formation of mind.* Cambridge: Harvard University Press.