Florida Journal of
Educational Research
Fall 1998, Vol. 38(1) 1 - 12

New Forms of Assessment in Reading: Capturing the Gist of Student Performance

Margaret Jorgensen
Martha-Anne McDevitt
Trudy Hensley

Educational Testing Service
Atlanta, Georgia

Sandra Wolfe

Polk County Public Schools
Winter Haven, Florida

ABSTRACT: The purpose of this paper is to share both the process and products of an innovative development project in reading comprehension assessment. This applied research project addressed the need for a valid and reliable measure that would capture credible and systematic evidence of reading comprehension abilities that teachers believed had been achieved by their students but were not being measured by traditional, standardized, norm-referenced multiple-choice tests. A notable challenge in this project was to develop an assessment with an engaging format and structure and to do so in a familiar context for reading that reflected sound instructional strategies.

The purpose of this paper is to share both the process and products of an innovative development project intended to create new forms of assessment in elementary reading comprehension. Driving this project was the observation by teachers in two Florida elementary schools that their linguistically diverse student population was more skilled in constructing meaning from text than was evident in their test scores. The goals of this project were to:

- Create reliable and valid assessments that closely mirrored sound instructional strategies
- Investigate the effect of alternative assessment formats on reading comprehension, and
- Engage teachers in the work of developing assessments so that they could serve as resident mentors for others who might wish to do similar work in reading or in other disciplines.

Setting the Stage--The Need for a Collaborative Effort

In October 1994, the Chapter 1 Rural Technical Assistance Center (R-TAC) that serves Florida was invited to give a presentation on the topic of new forms of assessment to the professional staff of Cypress Creek Elementary School in the Hillsborough County School System. This invitation was issued in response to a concern--expressed by members of both the faculty and the administration--that student learning was not adequately reflected in the standardized, multiple-choice, norm-referenced achievement test results reported annually. The teachers who worked with the students on a daily basis were convinced that important and meaningful learning was occurring and that this learning was not reflected in the test scores. At approximately the same time, a similar request for information on new forms of assessment came from Snively Elementary School, Polk County, Florida--a neighboring school system.

Not only were the needs and concerns expressed by each school similar, but so were many of their accomplishments, beliefs, and practices. Both Cypress Creek (a K-5 elementary school) and Snively (a K-6 elementary school) are exemplary schools. Both have received awards and recognition for innovations in instruction and in optimizing student achievement on selective and focused measures such as the "Florida Writes" direct assessment of writing. Both of these schools are actively engaged in site-based management. Both use technology to create dynamic, student-centered learning environments. Both have a significant population (50%) of students with limited English proficiency and both have large migrant populations (approximately 85%).

The initial inquiry from these two schools sparked a collaborative relationship involving the schools' faculties and administrations, the staff of the Rural Technical Assistance Center (R-TAC), and test developers at Educational Testing Service (ETS). The nature of the collaboration moved rapidly from one of information-sharing to one of product development using the resources and talents of each member of the collaborative. The goal of the collaborative was to produce assessments that were meaningful, credible, and informative for a variety of audiences and that would complement existing assessment information, thereby providing a full picture of students' capabilities with regard to reading comprehension.

The opportunity to work with schools committed to achievement at high levels for all students presented a unique opportunity to the R-TAC and ETS staff. We began by sharing with teachers and administration a brief overview of new forms of assessment. The examples shared were chosen to broaden the teachers' perspectives of what a test could look like. This was a deliberate attempt to encourage the teachers to use instructional strategies as a base for assessment design. This initial information sharing brought the members of the collaborative to a common understanding about the purpose, structure, limitations, and advantages of various forms of constructed-response (as opposed to only-choice) formats. It was critical to the credibility of the project that these new assessments sustain the attributes that traditional assessments typically use to demonstrate credibility. Thus, the elements of objectivity, reliability, and validity were essential to, and a primary directive in, all development work.

The initial conversations among the members of the collaborative focused on two key questions in the search for a better assessment tool in reading:

- What is quality work in reading at the grade levels of interest?
- How can evidence of reading comprehension that is consistent with our understanding of quality work in reading be captured?

Out of the intellectual struggles with these questions came solutions consistent with good practice in both teaching and assessment.

Answering the Key Questions

What Is Quality Work in Reading?

Task development began with a critical conversation focused on the question: What is quality work in reading at the grade levels of interest? Teachers were asked to identify the behaviors that an independent, competent reader should be able to demonstrate upon completion of the terminal grade level at each school, i.e., either fifth or sixth grade. From this benchmark, the notion was that the characteristics of the able reader at each grade level could be identified in a relatively straightforward way. The process hypothesis here was that once teachers had identified the benchmark behaviors that are characteristic of an able reader, conversations about how to capture evidence of those behaviors would lead to the development of assessments and scoring rubrics.

This task proved to be a difficult one for the members of the collaborative. Initially, their focus was on prerequisite skills such as print awareness (whether students knew word boundaries and conventions of print), discrete skills (recall of details and sequence, identification of the main idea, identification of fact versus opinion) and behaviors that suggested ability (make connections from the text to personal experiences, read for enjoyment, retell a story, select books of appropriate difficulty from the library). And, while these skills and behaviors are important during a student's development in reading, they do not fully describe the able reader and the integrative processes important to comprehension. Without a more complete description of how meaning is constructed by readers as they progress from novice to expert, the development of tasks to capture evidence about reading was both premature and inappropriate.

In order to expand their perspective of the reading process, the participants were asked to read a difficult passage and to engage in a think-aloud about the strategies they used to make sense--to capture the gist--of what they were reading. The strategies and processes reported in the think-alouds were consistent with the literature on expert readers (Kintsch & van Dijk, 1978; Afflerbach & Johnston, 1986) and included ignoring unimportant information, re-reading to gain understanding, making connections with prior knowledge, and using context and what is known about the structure of language and text to interpret the author's meaning. The insights teachers gained from this task influenced the format of the tasks that were designed and the standards of quality for judging student responses to the tasks.

Using the think-aloud experience, the identified comprehension processes, and teachers' prior understanding of what constitutes important evidence about reading, the team members were able to identify reader characteristics consistent with their curriculum and important to success in reading. Next, the team identified those reader characteristics for which credible, meaningful data were routinely collected by the classroom teachers and by school-wide testing. The product of this process was a matrix that could be used to document evidence of reader characteristics across a variety of tasks (matrix of evidence). After examining the matrix, the team determined that the area of constructing meaning (see Table 1) was not being adequately assessed by measures already in place. Therefore, capturing evidence of students' abilities to construct meaning became the focus of the task development.

How Can Evidence of Reading Comprehension Consistent with Our Understanding of Quality Work in Reading Be Captured?

The team's charge was to develop tasks with appropriate and engaging formats that would maximize the likelihood that students would be able to demonstrate comprehension of written text. The development of the new tasks was guided by assumptions that format and structure variables in other measures may have interfered with students' abilities to demonstrate their reading achievement (McDevitt, 1989). Further, it was important to examine those areas in which students did demonstrate success, such as the statewide writing assessment, and to use similar activities as a scaffold for demonstrating comprehension in reading. Based on these considerations, the following guidelines for task formatting were proposed:

- Avoid multiple-choice questions and, instead, use formats supportive of instruction.
- Limit the amount of writing required by having the students mark their answers in the text when appropriate, thereby reducing the confounding effects of writing skills.
- Build assessments that could be delivered both by paper-and-pencil and by computer to take advantage of the schools' technology.
- Structure the questions (both explicit and implicit) so that they could be answered or supported by information presented in the text, thereby potentially improving agreement among raters scoring student work.
- Restrict the text to primarily narrative text (stories) to match what teachers were teaching.
- Use complete works from children's literature in the development of tasks in order to support instructional practice.

Beginning the Hard Work

The development process began in January 1995, and the first phase was completed in May 1995. Two additional Polk County schools (Lena Vista and Scott Lake) sent observers to the development meetings.

Table 1 *Opportunities for Capturing Evidence about Reading Comprehension*

Reading Characteristics	Evidentiary behavior	Task description	Guided instruction	Text/Stimulus characteristics	Standards for scoring
		Constructs meaning through synthesis			
Understands relationships among important ideas in the text	Retells	Drama Oral response Written response Text marking Ability to answer questions	Advance organizers, and embedded questions provided	Story Two episodes All elements explicitly stated in text Passage length, 3.5 Flesch-Kincaid	Includes the elements of story or episode (plot, sequence of events, character motives, problem/solution, theme, lesson/goal, etc.) without extraneous details
	Identifies important hierarchical relationships	Ability to construct graphic of relationships Close: Connectors	Secondary cause/effect relationships provided	Cause/effect Explicit main idea Passage length, 5.5 Flesch-Kincaid	Identifies relationships among important ideas, with subordinates and superordinates correctly placed
Understands the gist of the text	Summarizes	Written response Text marking (deletion)		Cause/effect Compare/contrast Topic examples	Includes important ideas, given structure, and relationships among ideas, Excludes unimportant information
	Determines main idea	Written response Oral response Text marking			
Uses processing strategies	Reports strategies	Self-report Think-aloud			Rereads to confirm Self-corrects
	Displays strategies	Teacher observations and interviews			Uses prior knowledge
		Constructs meaning through analysis			
Understands explicit and implicit information (textbound)	Infers information	Ability to answer questions and complete story			Predicts outcomes
	Recalls information	Ability to answer questions and follow directions			Draws conclusions Determines fact/opinion, fantasy/reality
	Categorizes/ classifies	Ability to label groups and identify membership			
Understands conventions of print	Identifies words, sentences, etc.	Ability to use interviews and observations to answer questions			
	Uses titles, pictures, bold print, etc.	Ability to make predictions and construct questions			
Understands language cues	Uses semantic cues	Close Miscue analysis			
	Uses syntactic cues	Ability to answer questions			
	Uses graphic-phonemic cues	Various word analysis tasks			
		Reads beyond assignments			
	Reads independently for information and pleasure	Report of books read Self-report			
	Chooses appropriate books for skill and purpose	Interview Observation			

The development of the first generation of tasks was guided by a prototype (McDevitt, 1994) designed to reflect many of the components of a "Guided Reading-Thinking Activity" as described in Templeton (1995, pp. 257-271). For example, the story chosen was engaging. Advance organizers were included in order to focus the readers' attention on important ideas. Questions were asked at critical junctures throughout the story, and the nature of the questions supported reading as a problem-solving, active, and reflective experience. Further, the task itself required readers to document their thinking by marking evidence for their answers in the text.

The members of the collaborative agreed that the prototype story was very consistent with the types of materials used in their instruction, much more so than stimuli used in other assessment measures. Further, the members supported the development of questions, both explicit and inferential, that focused on important ideas, central to the story (Pearson & Fielding, 1991). It was hypothesized that both the format, which was familiar and user-friendly, and the nature of the questions (Pearson & Fielding, 1991; Irwin, 1991, pp. 179-193; Templeton, 1995), would contribute to more in-depth processing of the text, thereby improving students' performance in comprehension.

The text was arranged in a manner typical of a page layout in a storybook with minimal graphics. The text was divided into logical episodes or segments, which were preceded by advance organizers and then followed by questions pertinent to that part of the text.

By April 1995, the collaborative members had developed twelve reading assessment tasks. Ten of the tasks were modeled after the prototype (story with embedded questions). The other two tasks were expository texts about which students were asked to complete concept maps to assess their comprehension.

For those tasks modeled after the prototype, a common rubric was developed. This four-point rubric (1 to 4) was designed so as to give increasing credit for increasing levels of correctness. For example, a score of 3 was assigned to a response that included all of the target ideas necessary to answer the question but included some extraneous information, while a score of 4 was assigned to a response that included all of the target ideas without extraneous ideas. (Target ideas are those ideas [pieces of information], explicit in the test, that are determined to be necessary to answer the question.) For the tasks where students completed a concept map, rules unique to each task were developed.

Volunteers from the schools agreed to administer the draft assessments. The reservations of some volunteers suggested that they believed the tasks were too difficult for their students. Much to these volunteers' surprise, the students were able to do the work required without apparent frustration or stress. And many students reported that they enjoyed this work--certainly an atypical response to testing.

In May, the collaborative members scored the 1,147 student responses. Table 2 provides an overview of the twelve reading assessments by task title, target grade level, and number of students who performed each task.

Table 2 *Focus and Summary Data for the Reading Tasks*

Target grade level	n	M	SD	No. questions asked
		Carolina and the Caterpillar		
1	114	22.4	12.6	6 questions × 2 raters 48 possible points
		Creep-mouse		
5	55	33.1	11.5	7 questions × 2 raters 56 possible points
6	13	32.7	10.8	
Total	68	33.1	11.5	
		Disappearance of the Dinosaur		
5	11	32.7	14.2	6 questions × 2 raters 48 possible points
		Gregory the Terrible Eater		
3	48	15.1	9.6	6 questions × 2 raters 48 possible points
		Henrietta		
3	20	21.3	9.5	5 questions × 2 raters 48 possible points
4	90	24.8	12.1	
5	68	30.4	10.2	
Total	178	26.6	11.5	
		The Hippopotamus That Ate the Teacher		
2	111	14.3	12.8	5 questions × 2 raters 40 possible points
3	48	24.7	13.2	
4	41	33.3	9.9	
Total	200	20.7	14.5	
		Humans to the Rescue		
4	59	32.7	13	8 questions × 2 raters 64 possible points
5	40	33.5	15.2	
Total	99	33.0	13.9	
		A List		
2	36	23.3	13.1	6 questions × 2 raters 48 possible points
3	18	22.2	14.3	
Total	54	22.9	13.4	
		Just One Wish		
4	21	16.3	8.7	4 questions × 2 raters 32 possible points
5	8	6.0	4.0	
6	35	25.1	5.9	
Total	64	19.8	9.3	
		The Rainbow Fish		
2	51	16.2	12.5	6 questions × 2 raters 48 possible points
3	73	32.6	13.0	
Total	124	25.8	15.1	
		The Day Sweeney Ate the Worms		
5	68	66.1	28.0	13 questions × 2 raters 104 possible points
		The Wednesday Surprise		
3	124	28.8	16.1	9 questions × 2 raters 72 possible points
4	35	40.4	15.7	
5	18	21.0	11.6	
Total	177	30.3	16.5	

Note: The Disappearance of the Dinosaur and Humans to the Rescue represent expository text.

Sharing the Results

The importance of this project can be addressed from multiple perspectives. First, the impact of the task development process on teachers' understanding of reading and how reading comprehension should be taught is critically important. Second, the potential of the task design and content to document reading comprehension appears to be a viable complement to traditional reading tests. From yet another perspective, the positive model presented by this type of assessment has the potential to shift teachers' and students' thinking about what standardized testing can be. Each of these perspectives deserves further discussion.

Teachers' Understanding of Reading and How Reading Comprehension Should Be Taught

The impact of this work on teachers' thinking about reading comprehension, and about how to better create learning environments to support the construction of meaning from written text among diverse learners, is summarized by Sandra Wolfe, Principal of Snively Elementary School:

> The hard work began when ten teachers volunteered to work with ETS and the R-TAC staff to design performance-based assessments that would really capture credible evidence about what our students were learning. We knew we were teaching our thematic units effectively and that all students were learning tremendous amounts of good, rich, content-based information. But in the spring, our (standardized, norm-referenced) test scores did not reflect what our students were learning. We wanted credible evidence of learning!
>
> The journey was difficult. We alternated our time with ETS staff and time on our own as a group. Those times on our own were the most frustrating. We were so uncertain of whether or not we were right! Just like the students--we wanted to know immediately if we had understood the assignment correctly. Had we followed directions? Had we produced quality work?
>
> By the end of the year, we had produced twelve innovative performance-based reading tasks! We had crafted meaningful scoring guides that reflected increasing sophistication in reading, and we had tried out these tasks on students. Well, guess what? They worked, and the students liked them!
>
> For us, the "What's next?" question came quickly. We reflected on how much we had grown in terms of understanding. We recognized how valuable the products of our work had already been in making high quality work an expectation--a norm, not a distant goal in our school. Answering that critical question of "What is quality?" and then moving to the creation of systematic ways to document and judge quality have made this learning journey the most powerful that I, as a former teacher, have ever been on. Never again will we look at reading, the assessment of reading, or reading instruction in the same way. All students at our school will benefit from this work!

It is not easy to summarize or quantify the apparent impact of conversations about what quality work is. Nor is it easy to document how assessment development work brings about change in pedagogy, expectations about student learning, and instructional practices. However, it seems

reasonable to speculate that by keeping actively involved in assessment development, teachers will be pushed and pulled into continually improving the design of their instruction and assessments to accommodate an increasingly diverse student population. The notes and comments from teachers about how this professional development opportunity changed their classroom practices support this hypothesis.

Perhaps the most persuasive evidence that teachers valued this work was that they readily and enthusiastically chose to continue to participate in assessment development beyond the first year. In fact, many of the teachers from the collaborative volunteered their Saturdays to continue this work. That is, indeed, a strong index of support.

The Potential of These Assessments to Document Reading Comprehension Capabilities in Students

It is desirable, of course, to have the enthusiastic support of teachers and administrators in the development of new forms of standardized tests. It is also a plus if the students enjoy taking these new tests. However, the test's validity lies in the statistics. Can the student responses be scored reliably, and do the scores on these new performance tasks complement the evidence already collected about reading through well-accepted multiple-choice standardized tests?

Table 3 *Inter-rater Reliability Estimates*

Task title	Exact	Exact or adjacent
Carolina and the Caterpillar	74%	90%
Creep-mouse	80%	95%
The Disappearance of the Dinosaur	100%	100%
Gregory, The Terrible Eater	63%	87%
Henrietta	74%	91%
The Hippopotamus ... Teacher	84%	94%
Humans to the Rescue	92%	97%
A List	71%	91%
Just One Wish	88%	97%
The Rainbow Fish	76%	90%
The Day Sweeney Ate the Worms	88%	96%
The Wednesday Surprise	67%	87%

With regard to the first question--Can the student responses be scored reliably?--the answer is an unequivocal "yes." The percentages of exact agreements for these tasks are reported in Table 3.

Two independent judges had little, if any, difficulty in identifying the quality of student work and judging it in consistent ways. However, given the constraints of the response format, the high percentage of agreement is not a surprise. A practical advantage of the text-marking format was noted during the development of the training materials, as clear examples of each score point for all tasks were readily identified. In fact, the advantages of a common rubric and the abundance of clear examples provided the scoring director with the opportunity to conduct training on only two tasks. The remaining tasks were then scored without training specific to that task, as raters were expected to transfer what they had learned about the scoring parameters during training when scoring the remaining tasks. Although this is not an ideal practice, the time constraints imposed made more

lengthy training impossible. It is also the case that these scores were not going to be used to evaluate students, but rather to revise and improve the tasks.

Another way to gauge the impact of this work is to examine the correlations between performances on these new forms of assessments with scores derived from the traditional, norm-referenced, multiple-choice tests administered across the two participating school systems. If the tasks developed during this project capture evidence about students' ability to comprehend written text that is not captured by traditional, standardized, multiple-choice tests, then it is reasonable to expect that the correlations between the traditional and the innovative assessments would be low to moderate. If there were high positive correlations, there would be no evidence of complementary information. However, if the correlation coefficients were too low or negative, these tasks would not be considered credible sources of evidence of reading comprehension.

If, for the purpose of this discussion, moderate correlations are defined as .3 to .7, then all but one assessment task (Just One Wish) reported in Table 4 fall into that category with respect to the correlations. It is also important to note that the student population of both participating schools are highly transient, with a migrant student population of approximately 85%. This high percentage made it difficult to obtain large sample sizes of students having both norm-referenced test scores from spring 1994, and text marking scores from spring 1995. Thus some cells have small samples.

Table 4
Inter-score Correlation Coefficients

Task	Total	Sample Size			Correlations	Correlation	
		Vocabulary Score	Compre-hension Score	Task Rating		Vocabulary with Rating #1	Comprehension with Rating #2
Carolina and the Caterpillar	54	0	54	50	50	NA	.601
Creep-mouse	69	41	41	68	41	.521	.621
The Disappearance of the Dinosaur	20	0	0	11	NA	NA	NA
Gregory, the Terrible Eater	18	22	22	48	18	.266	.352
Humans to the Rescue							
The Wednesday Surprise	184	96	95	177	91	.527	.597
Just One Wish	71	33	33	64	33	.302	.203
A List	54	18	18	54	18	.411	.718
The Day Sweeney Ate the Worms	76	52	52	68	47	.359	.450
The Rainbow Fish	124	119	118	124	116	.708	.774
Henrietta	182	114	115	178	112	.328	.405
The Hippopotamus That Ate the Teacher	202	147	146	200	146	.590	.692

Note: NCE's were used for the standardized norm-referenced tests and raw scores for the new reading tasks.

It is also important to note that the score variability on the new assessment tasks is from 1 to 4. Although the distributions were fairly normal for each of the tasks (see Table 2), the restriction in

range of a four-point scale correlated with the Normal Curve Equivalent scale from either the Stanford Achievement Test (Hillsborough County) or the Comprehensive Test of Basic Skills (Polk County) does potentially impact the viability of the correlation analysis. Using Table 4 as a source of preliminary information, the majority of the new assessment tasks did yield low to moderate correlation coefficients between either the Vocabulary subtests or the Comprehension subtests. Thus, preliminary data support the notion that the collaborative members developed tasks that complement the data on reading comprehension as reported by the traditional, multiple-choice tests used in these two districts. Further research in this area will be conducted during the 1995-1996 school year using students' written responses to textually as well as scriptally implicit questions.

The range of students' performances within a given classroom seem to be consistent with the expectations of their classroom teacher. The range of performances across classes and across schools was often surprising, as was the pattern of scores on the same task across different grades. It is hoped that an increased number of students taking the tests, as field testing continues, will increase the *n* sufficiently so that the conclusions drawn from the data will be less subject to error. With larger sample sizes, it will be possible to benchmark the tasks for specific grade levels. This, in turn, will allow teachers to keep track of student progress.

This Type of Assessment Has the Potential to Shift Teachers' and Students' Thinking About What Standardized Testing Can Include.

The informal reports of student engagement by the teachers were notably positive. The teachers often, in fact, were surprised by the positive response of the students to these assessment tasks. Certainly the continuing commitment of the collaborative members to this work suggests that the participants, and the schools administration/faculty they represent, perceive the value and utility of different types of systematic assessments as effective tools to describe student learning. Whether or not the multidimensional approach to testing will be supported by the communities and by the school districts, however, is not yet clear.

Conclusion

The work continues, and the participating teachers are committed to developing and using assessments that more closely reflect ideal instruction and that enable students to demonstrate how they can think and what they can do. The assessment questions have served as models for questions used in reading instruction. The format, as well, has been integrated into reading activities across the content areas. Teachers are convinced that student learning has improved more rapidly than they would have expected because of this shift in instructional practices.

The development process has had a more lasting impact than any of the assessments themselves. The assessments are, in fact, less important than the transformations of the ways in which teachers and students work to build comprehension skills. Indeed, the improvement in standardized test scores for spring 1996 reported for Grade 4 at Snively Elementary School indicate that performance in traditional assessments may also be favorably influenced.

Authors' Note

Trudy Hensley, Director, Region 14, Comprehensive Assistance Center, Atlanta, Georgia; Margaret Jorgensen, Assistant Vice President, Iowa City, Iowa; Martha-Anne McDevitt, Associate Examiner, Atlanta, Georgia; Sandra Wolfe, Principal, John Snively Elementary School, Winter Haven, Florida.

To the principals, teachers, and students of Cypress Creek Elementary in Ruskin, Florida; Snively Elementary School in Winter Haven, Florida; Bartow Academy in Bartow, Florida; Lena Vista Elementary School in Auburndale, Florida; and Scott Lake Elementary School in Lakeland, Florida, go our heartfelt thanks for their time, their talents, their creativity and their courage. Their commitment to quality education for all students has stimulated and energized us all.

Correspondence concerning this article should be addressed to Margaret Jorgensen, ACT, Inc., 2201 North Dodge Street, PO Box 168, Iowa City, Iowa 52243-0168.

References

Afflerbach, P., & Johnston, P. (1986). What do expert readers do when the main idea is not explicit? In J. F. Baumann (Ed.), *Teaching main idea comprehension*, (pp. 49-72). Newark, DE: International Reading Association.

Irwin, J. W. (1991). *Teaching reading comprehension processes.* (2nd ed.) Boston: Allyn & Bacon.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of test comprehension and production. *Psychological Review, 85,* 363-394.

McDevitt, M. (1989). *Factors influencing main idea comprehension.* Unpublished manuscript.

McDevitt, M. (1994). *Children's construction of main ideas for expository text.* Unpublished doctoral dissertation, Emory University, Atlanta, GA.

Pearson, F., & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. L. Kamil, P. Mosenthal, and P. D. Pearson (Eds.), *Handbook of reading research*, (vol. 2), (pp. 815-860). New York: Longman.

Templeton, S. (1995). *Children's literacy, contexts for meaningful learning.* Boston: Houghton Mifflin.