

# Principles for literacy assessment

PETER JOHNSTON

The University at Albany-SUNY, Albany, New York, USA

PAULA COSTELLO

East Carolina University, Greenville, North Carolina, USA

In a “learning society” everyone will need to become, and remain, committed to learning. If assessment potentially represents the key to achieving this, it also currently represents the biggest single stumbling block. (Broadfoot, 2002, p. 6)

“What gets assessed is what gets taught” is a common assertion whose meaning is often underestimated. It is not just *what* gets assessed, but *how* it is assessed that has implications for what is learned. When a child who is asked the meaning of his report card grades responds, “If I knew that I’d be the teacher” he is saying something about the relationships of authority learned in the process of assessment. When a teacher wishes out loud that her faculty “could discuss retention and realistic expectations for grade levels without the nastiness and accusations,” she is also reporting on the relational aspect of assessment practices (Johnston, 2003, p. 90). Our goal in this article is to offer a framework for understanding literacy assessment that incorporates these dimensions and reminds us of the broader picture of literacy assessment of which we often lose sight.

## Literacy is a complex construct

Although we often think of literacy as a set of all-purpose skills and strategies to be learned, it is more complex, more local, more personal, and more social than that. Becoming literate involves developing identities, relationships, dispositions, and values as much as acquiring strategies for working with print (Brandt, 2001; Collins & Blot, 2003; Gee, 2000). Children becoming literate are being apprenticed into ways of living with people as much as with symbols. Consequently, literacy assessment must be grounded in current understandings of literacy and society (Johnson & Kress, 2003; Johnston, 1999). We have to consider what kind of

literacy might benefit individuals, what kind of literate society we aspire to, and what assessment might best serve those ends.

For example, what kind of literacy assessment will enable children to live in and contribute to an increasingly democratic society? Democracy has to do with “the way persons attend to one another, care for one another, and interact with one another...[and] the capacity to look at things as though they could be otherwise” (Greene, 1985, p. 3), and citizens who “have the convictions and enthusiasms of their own responses, yet...are willing to keep an open mind about alternate points of view, and...to negotiate meanings and actions that respect both individual diversity and community needs” (Pradl, 1996, pp. 11–12). In other words, our literacy assessment practices must foster a literate disposition towards *reciprocity* (Carr & Claxton, 2002); that is, “a willingness to engage in joint learning tasks, to express uncertainties and ask questions, to take a variety of roles in joint learning enterprises and to take others’ purposes and perspectives into account” (p. 16).

What might such assessment look like? The National Educational Monitoring Project (NEMP) in New Zealand is charged with taking stock of the nation’s progress in educating a literate society. To this end, the NEMP includes items such as providing a group of children with a set of books from which they, as a class library committee, must make their best selection. Students individually justify their choices to the group before the group negotiates and justifies the final selection. The negotiation has a time limit and is videotaped for analysis of reading and literate interactions (Flockton & Crooks, 1996). This item requires children to evaluate the qualities of texts, take a stance, make persuasive arguments, actively listen, and negotiate a collective position—all independent and interdependent literate practices central to democratic classroom and society. The item reflects and encourages an individual and mutual disposition toward reciprocity, a foundation for a democratic literacy.

Literacy has complications that assessment must deal with. Not only is literacy complex and social but also the literate demands of the world keep changing with exponential acceleration. The apparent boundaries between spoken and written words and their conventions have been obliterated by instant messaging, book tapes, cell-phone text messaging, speech translation software, interactive hypertext, and the facility with which text and image (moving or still) are fused. Literate demands are changing so rapidly that we can’t predict with certainty what kindergartners will face in adulthood. We do know

however, that they will need to be resilient learners (Carr & Claxton, 2002) to maintain their literate development in the face of the increasingly rapid transformations of literacy in their communities.

Because “what is assessed is taught,” literacy assessment should reflect and encourage resilience—a disposition to focus on learning when the going gets tough, to quickly recover from setbacks, and to adapt. Its opposite is brittleness—the disposition to avoid challenging tasks and to shift into ego-defensive behaviors when learning is difficult. A brittle learner believes that having difficulty with a literate task reveals a lack of “ability.” A brittle disposition in children prior to first grade negatively predicts word recognition in grades 1 and 2, and is a better predictor than assessments of phonological awareness (Niemi & Poskiparta, 2002). This negative effect on learning is amplified by the pressures of competitive and overly difficult situations, particularly where ability is the primary emphasis. These are exactly the contexts produced by current testing practices.

Resilience can be assessed. For example, teachers can collect specific examples of resilience with quotes and artifacts to produce documented narratives (Carr & Claxton, 2002) for later review with the student and other stakeholders (see also Himley & Carini, 2000). In fact, the process of generating such assessment narratives will foster a resilient literate disposition (Johnston, 2004).

We begin with these uncommon examples of literacy assessment to suggest that, although assessing literacy in its complexity can be challenging, it is possible. It is also important. Failure to keep our attention on the bigger picture might not be a problem except that, intended or not, literacy assessment instruments define literacy within the assessment activity and, particularly when the stakes are high, within instruction (Smith, 1991). The higher the stakes, the more necessary it is that assessments reflect the breadth of literacy. Alas, most assessment practices, particularly testing practices, oversample narrow aspects of literacy, such as sound-symbol knowledge (Stallman & Pearson, 1991), and undersample other aspects such as writing, any media beyond print on paper, and ways of framing texts and literacy, such as the critical literacies necessary for managing the coercive pressures of literacy.

The more an assessment focuses on a narrow sample of literate behavior, as happens in individual tests, the more undersampling occurs. Literacy assessments distorted in this way affect instruction in many subtle ways. For example, the extensive use of pencil-and-paper state tests has forced many teachers to decrease instructional use of computers, particularly

for writing. This problem is most damaging in urban and poor-performing schools (Russell & Abrams, 2004). The tests simultaneously risk underestimating the writing competence of students used to writing on computers, while reducing the likelihood of students not familiar with computer writing to ever become so.

## Assessment is a social practice

Assessment is a social practice that involves noticing, representing, and responding to children's literate behaviors, rendering them meaningful for particular purposes and audiences (Johnston & Rogers, 2001). Teacher feedback to students on their literate behavior is assessment just as much as is grading students' work, classifying students as handicapped, certifying students as being "above grade level," or establishing a school as "in need of improvement" (Black & Wiliam, 1998a; Johnston, 1993). *Testing* is a subset of assessment practices in which children's literate behavior is elicited in more controlled conditions.

Although assessment often is viewed as a technical matter of developing accurate measuring instruments, it is more centrally a set of social practices in which various tools are used for various purposes. For example, leveled books can be used as part of teaching in order to monitor children's early reading growth without the use of tests. Some books can even be kept aside specifically for assessment. The same procedure could also be used as part of holding teachers accountable for children's progress (Paris, 2002). However, this is a very different social practice and would invite greater concern about the measurement precision of the "levels" and different social action. For example, teachers would be more likely to use the assessment books for instruction and to focus the curriculum on the accuracy of word reading.

Although the instrument is the same, it has different meaning in the different social practice. In the accountability context, we worry more about the measurement qualities of the instrument in order to be fair. Fairness in the teaching context is more about ensuring that children are developing adequately, focusing instruction, and ensuring that the discourse of "levels" does not dominate the children's interactions and self-assessments. Paradoxically, though we worry more about the psychometric properties of an instrument in the accountability context, the social properties of the *use* of the instrument, such as the defensiveness it might induce, or

the constriction of the curriculum, can be of far more significance.

With the realization that assessments are social practices has come the awareness that the validity of an assessment instrument cannot be established outside of its consequences in use (Messick, 1994; Moss, 1998). Literacy assessment practices affect the constructs used to organize teaching practice and to represent children (Johnston, 1997; Moss). This is especially powerful when tests are used for purposes that attach high stakes such as teacher salaries, student retention, graduation, or classification.

Although there are occasional studies claiming that high-stakes testing has no negative effects, or even some positive effects on children's learning, there are many more studies showing the opposite and with greater specificity. For example, high-stakes accountability testing has consistently been demonstrated to undermine teaching and learning (Allington & McGill-Franzen, 1995; Morrison & Joan, 2002; Rex & Nelson, 2004; Smith, 1991; Smith & Rottenberg, 1991) particularly for lower achieving students (Harlen & Crick, 2003). It restricts the literacy curriculum, thus defeating the original intention to improve literacy learning. Teachers under threat drop from the curriculum complex literacy practices involving, for example, multimedia, research, and role-play, and at the same time their learning community is disrupted (Rex & Nelson). Increasing accountability pressure on teachers is counterproductive, especially when teachers already have an internal accountability system. It results instead in "escalating teacher outrage, diminishing moral [sic], and the exiting of committed teachers...from teaching" (Rex & Nelson, p. 1324).

The dictum "first do no harm" has become part of validity in theory, though rarely in assessment practice. Indeed, although high-stakes testing has lately been supported by arguments that it will reduce literacy achievement differences associated with race and poverty, there is evidence that the long-term effect of such testing is to create a curriculum that extends stratification rather than reducing it (Darling-Hammond, 2004; McNeil, 2000).

## Individual and institutional learning

Literacy assessment is part of a larger project to educate children both for their immediate and long-term benefit and for the evolution of society. The implication of this is that literacy assessment must be

grounded in current understandings of individual and institutional learning. There are two general kinds of assessment—summative and formative. Summative assessments are the backward-looking assessments *of* learning, the tests we most commonly think of that summarize or judge performance as in educational monitoring, teacher and student accountability testing, and certification (Black & Wiliam, 1998a). These have not been overtly associated with current understandings of individual or institutional learning.

Indeed, the theories of learning underlying psychometric practices have largely been implicit, individualistic, and behavioristic (Shepard, 1991). For example, current accountability testing, driven by psychometrics, is based on rewarding and punishing students, teachers, and school systems. The evidence so far is that, rather than accomplishing the intended learning, these practices shift participants' goals toward avoidance of punishment, which thwarts the goal of improving the quality of literacy learning for all students and particularly for historically low-achieving students (McNeil, 2000).

Formative assessment, or assessment *for* learning, is the forward-looking assessment that occurs in the process of learning, the feedback the teacher provides to the student, and the nature of the feedback matters (Crooks, 1988). For example, rather than praise or grades, comments improve performance, though praise keeps students thinking they are doing well (Black & Wiliam, 1998a). Feedback that focuses attention on traits such as ability, smartness, or goodness, undermines resilience (Dweck, 1999).

But the *process* of formative assessment is also critical. For example, the most common assessment practices associated with comprehension involve asking for retellings or asking questions to which teachers already know the answers. These interactional patterns teach children about how literacy is done and how authority is organized (Johnston, Jiron, & Day, 2001; Nystrand, Gamoran, Kachur, & Prendergast, 1997). Arranging for children to ask the questions and selectively discuss them can provide more interesting information regarding children's understanding, while simultaneously socializing them into productive literacy practices and identities (Comeyras, 1995).

Formative assessment is specifically directed toward affecting learning. Its validity depends on its ability to do so (Crooks, 2001). This means that the validity of formative assessment rests on factors not normally considered in discussions of validity, such as trust and sensitivity, the social supports, and motivations of the classroom. Task factors will be important,

such as the nature and difficulty of the task, its personal and external relevance, the articulation of task features, and performance criteria. Each of these will affect the development of self-assessment. The nature and timing of feedback will be important. But because human interactions are structured around who the participants think they are and what they think they are doing, teachers' understanding of such things as literate practice, how children learn, and cultural difference will also be important, as will their social imagination and insight on conceptual confusions.

While this is true of formative assessments, summative assessment practices affect learning too. Some, such as accountability testing, do so deliberately. Consequently, to be valid, *all* assessment practices should be grounded in current and consistent understandings of learning, including the above factors. Both summative and formative assessments participate in socializing children's and teachers' self-assessments, with implications for control of learning and the management of self-assessment to serve learning goals.

Basing assessment on current understandings about learning does not simply negate principles of psychometrics. For example, neo-Piagetian theories of learning view the process of confronting and resolving discrepancies as a primary vehicle for learning (Schaffer, 1996; Tudge & Rogoff, 1989). A self-extending literacy learning system requires children to attend to discrepancies between cue systems, for example (Clay, 1991). In a similar way, learning communities require disjunctures, such as between minority and mainstream performance, to stimulate learning. However, as with formative assessment, the independent sources of information providing the conflict must be trusted, and measurement principles can help provide the grounds for this. The context in which such discrepancies are presented affects what is learned. The assessment activity must enable productive *engagement* of the disjunctures and foster productive use of data.

Thinking about assessment in terms of individual and institutional learning can change the way we value technical characteristics of assessment. For example, consider the role of consistent agreement among examiners (reliability). Complex authentic assessment items such as those used in the NEMP often reduce reliability (Shavelson, Baxter, & Pine, 1992). Weighty assessment practices like sorting and certifying students demand practices that ensure agreement—the higher the stakes, the more important this agreement.

Disagreements in this context are viewed as “measurement error,” which leads to a reduction of

complex authentic items. By contrast, in low-stakes and more formative assessment, disagreement among teachers about the meaning of particular documentation, such as portfolios, can open an important learning space by inviting discussions that lead to improvements in instruction and assessment itself. Indeed, this negotiation of values, qualities, and purposes is the most productive part of standards-based or performance-based assessments (Falk, 2001; Johnston, 1989; Moss & Schutz, 2001; Sadler, 1987). Complex and problematic examples provoke the most productive teaching-learning conversations. In other words, when the stakes are low, the less reliable the assessment is—to a point—the more likely it is to produce new learning and innovation in teaching. Because the validity of an assessment rests partly upon its consequences, improving teaching increases the validity of the assessment. In this context, imperfect reliability, contrary to psychometric theory, can increase validity.

As a concrete example, consider the NEMP test item mentioned at the beginning of this article in which children evaluate books individually and collaboratively as a library committee. The item and instrument are possible because NEMP uses a light matrix sample. Different children take a different selection of items; nationally, only a sample of children takes any items at all. The sampling system is possible because the emphasis is on the performance of the system, not of individual schools, teachers, or children. The instrument provides system information without raising individual or organizational defenses (Argyris, 1990).

Aggregate performance is published and analyzed by kind and size of school, minority percentage, community size, socioeconomic status, ethnicity, and gender, but direct institutional comparisons cannot be made. Test items are also published to reduce emphasis on abstracted numerical comparisons. The four-year assessment cycle allows time for both the construction of complex assessments and productive institutional and societal responses. At the same time, each administration of the assessment requires training a group of teachers to reliably administer the assessment. Teachers involved in the training report that it is an exceptional form of professional development that influences their own assessment and teaching competence, and that they pass this competence on to others (Gilmore, 2002).

## Minds in society

Children's thinking evolves from the discourses in which they are immersed. So, for example, the

ways children assess themselves as literate individuals will reflect the discourse of classroom assessment practices. Consider Henry (all names are pseudonyms), for example, a fourth-grade student who describes himself as a writer (Johnston et al., 2001). Though he says writing takes him a little longer than some, he notes that he has a journal with lots of entries and can borrow ideas from other authors, among whom he includes peers whose feedback and suggestions he values. He talks about their writing in terms of the ways they can affect him as a reader. He enjoys reading, and if he wanted to learn about another person as a reader, he would ask about favorite and current books and authors.

Indeed, he describes peers first in terms of their reading interests (topic, author, genre, difficulty) and then, matter-of-factly, their reading speed. He is confident that he makes important contributions to book discussions, but he also feels he benefits from hearing other students' experiences and interpretations. He has learned to manage these discussions to maximize this learning. In his research efforts he has encountered disagreements among authors, which he ascribes to one of them not "doing his homework," and he resolves them by consulting more sources (print, personal, and electronic). Henry has a strong sense of agency and uniqueness in his literate practice, which is an important part of who he feels he is. He recognizes a range of sources of authority and that none is beyond critique. When his teacher describes Henry's literate development, it is in terms of details of his interests and engagements, what he has accomplished, how he approaches literate activities, and what he is beginning to do collaboratively or with assistance.

Henry's self-assessment, his interpretation and representation of himself as a literate person, reflects the literate practices and values of his classroom. In a different discourse community, his and his teachers' assessments could have focused more centrally on his decoding skills, what he is *unable* to do, or on his normative standing. The test used by his school district does provide a numerical quantity to represent the amount of his literacy and places him in the lower quarter of his class. But this particular teacher in this particular school and district finds that representation of little significance, and it does not enter the discourse of the classroom. Another teacher in another discursive community in which the pressures and goals are different would likely represent the child's literacy development differently.

Indeed, teachers in districts more concerned with accountability pressures tend to describe children's literacy development with less detail, with less

attention to the child's interests, and with more distancing language (Johnston, Afflerbach, & Weiss, 1993). In a similar manner, the pressures of standards assessment change not only the representations but also the relationships among teacher and students, making them more authoritarian (Deci, Siegel, Ryan, Koestner, & Kauffman, 1982), a relationship that is part of the literacy that is acquired.

A corollary of the "mind in society" principle is that literate development is constructed. Mandy, for example, in the same grade in another school district, feels that she is a good writer because she "writes fast" and feels that she will get an "excellent" on her report card for writing with a comment that she "has behaved and she is nice to other classmates." She feels that the good readers are recognizable because they "are quiet and they just listen...and they get chapter books." However, she does not think that conversations between writers are good because they would result in other writers taking ideas and having the same stories and because feelings might get hurt.

Mandy's conception of literacy foregrounds convention, conformity, speed, and individualism (Johnston et al., 2001). Rather than acquiring similar amounts of literacy, as their test scores might suggest, Henry and Mandy have acquired different literacies. Literate development is not a matter of acquiring a series of stepping stones in a particular order. First graders are quite capable of acquiring knowledge of letters and sounds and other print conventions as part of developing a critical literacy. The conventions, however, will mean something different when acquired as part of different literacies. The fact that there are predictable sequences of development is as much a feature of our assessment and curricular imperatives as it is a feature of a natural sequence of literate subskills, or of biological or other potentials and limitations.

## Representation and interpretation

Assessment practices are always representational and interpretive. A teacher, an administrator, and a parent are likely to make different sense of a child's literate behavior both because they bring different histories to the assessment and because they often have different goals as part of different, if overlapping, social practices. Even a test score (a particular choice of representation) will mean different things to them. Each assessment practice is associated with distinct ways of using language that influence the interpretations made (Fairclough, 1992; Gee, 1996). A

school psychologist or a speech therapist can tilt the representational language of a committee on the handicapped toward "learning disabled" or "language delayed" on the basis of the same evidence (Rueda & Mercer, 1985). A single teacher can bring different discourses to representing different children depending on the way the child has been categorized, and these representations have consequences for children's understandings of literacy, themselves, and one another as literate individuals (Arya, 2003; Johnston et al., 2001).

Representational practices in assessment perpetuate the wider cultural discourses. If our discourse offers a category called "reading disabled," then we will find assessment tools to identify members of the category and an appealing narrative of "services" and "support" (McDermott, 1993). The representational language of trait and deficit (Johnston, 1993; Mehan, 1993) within which learning narratives are set offers children, teachers, parents, and other community members problematic identities and dispositions. Once "identified," children remain caught in the problematic discursive web, partly because the problem is represented as a trait of the child rather than as in the instructional environment, partly because the identification process groups children together who share common identifications, and partly because the child is moved to a system that specializes in children's problems that often emphasizes different understandings about literacy learning (Allington & McGill-Franzen, 1989).

Although we might worry about the nature of the categories, which are surely important, the practice is about more than that. As Yalom (1989) pointed out, "If we relate to people believing that we can categorize them, we will neither identify nor nurture...the vital parts of the other that transcend category" (cited in Greenberg & Williams, 2002, p. 107). This is evident in casual transformations such as "He's a two, borderline three, right now and we hope that this enrichment program will put him over the edge" (Baudanza, 2001, p. 8).

## Primacy of teachers' assessment practices

No instrument or assessment practice can overcome the fact that the teacher is the primary agent of assessment (International Reading Association and National Council of Teachers of English Joint Task Force on Assessment, 1994). The bulk of literacy assessment occurs moment by moment as part of the

activity of teaching (Black & Wiliam, 1998b; Crooks, 1988; Johnston, 1989). Consider an example. A teacher was observed introducing to a student a predictable book with the pattern "Grandpa is [verb—e.g., sitting]." The last page was "Grandpa is snoring," at which the child laughed and said that his grandpa snores too. However, when he read the book he read that page as "Grandpa is so funny." The teacher prompted the child to recall what his grandpa does and then prompted a rereading. The child reread, hesitated before *snoring*, and read it correctly.

But why *that* prompt or teaching strategy? Why not ask the child to read with his finger to emphasize the mismatch between the number of words spoken and in print? Because the teacher hypothesizes, based on her ongoing assessment of the child, that he thinks *so funny* is one word. Pointing would not prompt rethinking because he would still have a one-to-one match and an initial letter match. Why not simply provide accuracy feedback? Because, she hypothesizes, that the process through which the child solves the problem himself will help build a sense of literate agency. Her feedback is based on a theory of learning more than a notion of performance.

The essence of formative assessment is noticing details of literate behavior, imagining what they mean from the child's perspective, knowing what the child knows and can do, and knowing how to arrange for that knowledge and competence to be displayed, engaged, and extended. This requires a "sensitive observer" (Clay, 1993) or "kidwatcher" (Goodman, 1978), a teacher who is "present" in the classroom—focused and receptive to noticing the children's literate behavior (Rodgers, 2002). A child's acquisition of a "reading disabled" classification (and identity) begins with the teacher's assessment, and teachers who notice less about children's literacy development refer more children to be classified than do those who notice more (Broikou, 1992). The more detailed teachers' knowledge of children's literate development, the more agency they appear to feel with respect to solving literacy learning problems.

Formative assessment requires not only noticing and making productive sense of the literate behaviors that occur but also arranging classroom literacy practices that encourage children to act in literate ways and that make their literate learning visible and audible. A child explaining how she figured out a word is not only providing this information for the teacher but also spinning an agentic narrative of her own literate competence. She is building a productive self-assessment and literate identity (Johnston, 2004).

If a classroom is arranged so that children routinely engage in literate activities that provide man-

ageable challenges and talk about the process and experience of their literate practice, assessment information is available to the teacher and, simultaneously, strategic information is available for the students. *Play* is a particularly rich context for the display of young children's understanding of how literate practices work (Roskos & Neuman, 1993; Teale, 1991). In a similar way, *collaboration* demands an externalization of shared thinking, which also provides an excellent source of information.

To the extent that formative assessment is a technical matter, the "instrument" is the teacher and his or her mind and its social and textual supports. Improving performance on summative assessments requires improving formative assessment. There is research that suggests how to do this, but it also suggests that change will be slow because the practices assume active involvement on the part of students as well as changes in the ways teachers understand students, themselves, and what they are trying to accomplish (Black & Wiliam, 1998b). These changes are strongly resisted by societal assessment discourses and their sedimentation in teachers' own subjectivities, as we discuss presently.

## Literacy assessment and context

Literacy is somewhat local in that people engage in literate practices differently in different contexts. Different tools and social contexts invoke different strategies and ways of thinking. Common assessment practices do not recognize this fact; instead they assume that performance on a particular task in a testing context is representative of all literate contexts. But children perform differently, for example, in more meaningful or authentic activities. The Primary Language Record (PLR) (Barrs, Ellis, Hester, & Thomas, 1989), an early literacy assessment instrument, requires the assessment community (teachers, families, administrators, and students) to recognize (and document) performance in different contexts including "collaborative reading and writing activities," "play," "dramatic play," and "drama and storying" across different social groups that include "pair," "small group," and "child with adult" (p. 38). It draws attention to what a child can do independently and with different kinds of support.

Assessing children's literate learning requires attending not only to what they know and do but also at least as much to the context in which they know and do. Indeed, as the PLR manual notes, "progress or lack of progress should always be seen in relation to the adequacy of the context" (p. 18). When a

child appears to be unsuccessful at literate endeavors, we want to know the circumstances in which this happens. Such circumstances include the extent to which literate practices and the logic of participation are made visible in the classroom and valued as purposeful social activities, the extent to which materials are relevant and accessible, and the extent to which classroom discourse is supportive, specific, reflective, nonjudgmental, and values problem solving (Allington & Johnston, 2002; Johnston & Rogers, 2001; Pressley, Allington, Wharton-MacDonald, Collins-Block, & Morrow, 2001).

Shifting the focus of assessment away from the isolated mind to the mind in a social context has begun to be recognized in the assessment of reading disabilities. For example, Clay (1987) proposed that labeling a child as reading disabled is premature without first eliminating the possibility that the child's progress is a result of poorly configured instruction. The assessment strategy of providing the best instructional intervention we can muster has proven effective in eliminating the need to classify most children (Scanlon, Vellutino, Small, & Fanuele, 2000). However, because this strategy remains in a discourse that expects individual disabilities, the handful of children who remain unsuccessful become viewed as bona fide "disabled," or "treatment resisters" (Torgeson, 2000). This need not happen. Indeed, Smith and her colleagues (Smith, 1997) rejected that discourse. Instead of locating the problem in the child, they entertained the possibility that their intervention might still be insufficiently responsive. Through collaborative self-assessment using videotapes, they refined their intervention and produced the desired acceleration in literate learning, removing the need to classify even these students. This concept of attending to the child in the learning context might be applied to large-scale assessments too. Teachers and schools do not operate in a vacuum.

## Assessment discourses distribute power

Assessment discourses distribute and sustain power relationships. For example, formative assessments, while grounded in current understandings of learning, are not taken seriously as a form of assessment (Black & Wiliam, 1998a). They are referred to as "informal," as opposed to the more authoritative "formal" assessments. There are probably many reasons for their lack of institutional power aside from the fact that they don't always involve a textual record or arti-

fact such as running records, documented events, or writing samples. They are the purview of teachers, mostly women, and they are normally not in the language of mathematics. When brought to a Committee on Special Education meeting, these assessments are easily trumped by the tests of the school psychologist.

Rogers (2003) showed how a mother, vehemently committed to protecting her daughter from assignment to special education, is reduced to passive acceptance by an assessment discourse that invokes subjectivities from her own unsuccessful history in schooled literacy. Rogers also showed how the discursive context induces this passivity just as well in those with highly successful histories of schooled literacy. The normative discourse of testing provides a powerful tool for asserting symbolic domination and intimidation of students, teachers, and parents (Bourdieu, 1991; Fennimore, 2000; Rogers, 2002). When an adult basic education student at the end of a reading lesson asks timidly, "Did I read this good?" (Rogers, 2002), she demonstrates the internalization of an oppressive assessment discourse.

It is possible to design assessment practices to alter these power arrangements. To return to the PLR, the manual describes specific ways for reducing power differences in assessment conferences with children and families. The form of the assessment also insists that members of the learning community focus attention on the child's assets and their instructional context. Because it directs attention toward differences in performance in a range of contexts and on a range of dimensions, it resists narrow and debilitating ability interpretations. At the same time it provides a language that represents literacy as centrally involving identity and engagement in practice, describing a child's development as a reader and a language user and implying a dimension of agency.

However, breaking free of more limiting assessment discourses is increasingly difficult as these discourses saturate a wider array of media. Constant reminders in the newspaper and reports from school are now supplemented through the Internet. Parents going to the Web are encouraged to obtain reading tests that they can use with their child. Like any advertising, these tests create a need and then direct parents to purchase the remedial instruction on the basis of the normative assessment and the "latest brain research" (Learning, 2002) to fulfill the need. At the same site, parents learn of the routinely massive company growth rate, its even better prospects following federal No Child Left Behind regulations (2002), and how they can profit through investment (Johnston & Rogers, 2001). By both reflecting and enforcing traditions of literate practice (including



who gets to participate in what ways and in which media), assessment practices stabilize the literate society, limiting social change and adaptability.

## Clashes in practices

Literacy assessment consumes resources, so there is a constant search for multipurpose assessments. However, each new function often has different demands, requiring difficult trade-offs and bringing different discourses. Recall the NEMP assessment described earlier. Many of the features of the NEMP were once part of the National Assessment of Educational Progress (NAEP in the United States). However, political pressures have changed the timing of the NAEP to a two-year cycle, increasing pressure for simpler computerized responses. The sampling structure has changed to enable state-by-state comparisons, and state performance has become pegged to federal funding through the No Child Left Behind legislation (2002), thus increasing the assessment stakes. These changes add up to a change in the nature of the assessment activity from educational monitoring for productive curricular conversations to instrumental control of literacy teaching and learning. This is a different assessment practice, grounded in different views of learning and literacy.

The clash of these different discourses is common in school systems as formative and summative functions are forced together, often catching teachers in the middle (Delandshere, 2001; Hill, 2004). As with the earlier example of using leveled books for accountability practices, the higher stakes assessment will generally subvert the lower stakes practice. However, it is possible to have consistency among school literacy curriculum and assessment practices.

The PLR, described earlier, was developed in London for literacy assessment in multicultural/multilingual inner-city communities. It represents a complex, contextual, and social view of literacy learning and assessment practice that involves teacher, student, and parent in collaboratively documenting the child's literacy development over time. It was deliberately designed to inform and support teaching, students, and family literacies through clear documentation and the *process* of that documentation—the assessment *activity*. Although it is a “record,” its developers took seriously the educative, communicative, and relational dimensions of assessment practice. In systematic interviews, parents describe the child's home literacy and must agree on what is recorded. Because interview topics include “opportunities that might be possible for writing at

home and whether the child chooses to write” (Barrs, Ellis, Hester, & Thomas, 1989, p. 16), parents simultaneously learn about possible ways to expand family literacy practices.

The representation of the child is centrally focused on documentation of what the child does and how the child does it and understands it. In context, though, it also includes numerical ratings for aggregation at the institution level and to complement the descriptive detail. Serious professional development is required for a complex assessment system like the PLR. But that has not prevented its successful adoption (Falk, 1998; Falk & Darling-Hammond, 1993). Implementation is not expected to occur overnight, and it is recommended that teachers begin by selecting a small group of students to document, expanding the group as expertise develops.

However, much of the professional development is built into the process of the assessment. In order to obtain reliable ratings, participants in the assessment community (teachers, administrators, parent representatives) regularly gather to compare their analyses of one another's assessments. The discussion around cases of disagreement is productive in clarifying the need for recording detail and the bases for judgment. The public nature of these discussions keeps teachers responsible for their assessments and requires a measure of courage. Because the assessment requires a range of literacy learning contexts and particular kinds of evidence, it helps teachers to structure their classroom practice.

We provide this example to show that more common approaches to assessment should not be thought of as “givens” that merely need tweaking. This assessment holds very different assumptions from the more standard views and has very different consequences. For example, the assumption behind current accountability testing is that schools as organizations, and the individuals within them, are not only unable to monitor their own performance but also are unlikely to provide the best instruction they can unless forced to do so annually through rewards and punishments. The successful use of the PLR suggests that this assumption, at least in some contexts, is not tenable. Instead, we might sensibly ask, “Under what circumstances can organizations and individuals productively monitor their teaching and learning as part of improving literacy learning?”

Darling-Hammond (2004), examining successful examples of assessment-driven reforms, provided some answers, concluding that consistency in assessment and curricular imperatives across the institutional learning community is essential. Other critical properties that the system provides, in a time-

ly way, included sophisticated information that is consistent with current understandings of learning and relevant for teaching individual students. Successful assessment systems also provide information about the qualities of students' learning opportunities (the context of learning), develop productive teacher-student relationships, and are able to "leverage continuous change and improvement" through a focus on teacher quality and learning (p. 1078). She noted that relatively low stakes and consistency among the assessment and curricular imperatives are important and that institutional size is not trivial. Although Darling-Hammond focused on the testing context privileged in the United States, these emphases are exactly the design features of the PLR. This is a description of the PLR.

## Final comment

Assessment always (a) is representational and interpretive; (b) is a dynamic part of ongoing, goal-directed social activities and societal discourses; (c) reflects and imposes particular values, beliefs, relationships, and ways of being literate; and thus (d) has consequences for individuals' and communities' understandings of themselves and one another, as well as for the kinds of individuals and communities they will become. If the accelerating shifts in society will require everyone "to become, and remain, committed to learning," (Broadfoot, 2002, p. 6) and to acquire literacies that are more flexible and open, more resilient and self-directed, and more collaborative in a culturally and linguistically diverse context (Kalantzis, Cope, & Harvey, 2003), they will need to be socialized into a literacy that makes this possible, and our assessment systems are part of that socialization.

This means that learning must form the basis of our assessment practice. Current understandings show that the ability to guide and monitor one's own learning is essential to this project (Crooks, 2001). Focusing on learning in this way might incidentally accomplish other shorter term goals. For example, creating classrooms in which assessment practices socialize children into self-regulated literacy learning not only serves students' development as learners but also develops their literate achievement (Harlen & Crick, 2003; McDonald & Boud, 2003). The same principles almost certainly apply to teachers as individuals and as institutional communities. Indeed, if we are to have consistency among assessment and curricular imperatives within schools, the consistency should apply to the processes as well as the content. If literacy assessment is to serve literacy learners and society, then

it has to be grounded in processes that reflect current understandings of learning, literacy, and society. It also has to remain open to evolution in both literacy and assessment, which at the very least means encouraging some diversity in assessment practice.

Nearly a decade ago, Shepard and her colleagues interviewed officials from state departments across the United States and concluded that more complex and authentic forms of literacy assessment were developing and that the previous excesses and problems of assessing children, particularly young children, for high-stakes purposes like accountability and retention were largely gone (Shepard, Taylor, & Kagan, 1996). The opposite is now true, a development that has everything to do with politics and relatively little to do with research (Allington, 2002; Allington & Woodside-Jiron, 1999; Johnson & Kress, 2003; Wixson & Pearson, 1998).

Indeed, the United States has currently reached the highest volume of testing and the highest stakes testing in its history. We are reminded of a definition of fanaticism as the act of redoubling one's efforts while having forgotten what one is fighting for (de Toqueville, cited in Claxton, 1999, p. 281). Although this article is in the service of "theory and research into practice," we must not pretend that literacy assessment can be improved by simple application of either. At the very least our theory in practice has to include the fact that changing assessment practices is about changing societal discourses regarding children, literacy, and education, with all the values, relationships, identities, and resources that entails.

**PETER JOHNSTON** is a professor of education at the University at Albany-SUNY. He chaired the International Reading Association/National Council of Teachers of English Joint Task Force on Assessment that produced the position monograph *Standards for the Assessment of Reading and Writing*. Besides his interest in literacy assessment, his research investigates the consequences of teaching practices for the kinds of literacy children acquire. He can be contacted at ED309, The University at Albany, Albany, NY 12222, USA, or by e-mail at pjohnston@uamail.Albany.edu.

**PAULA COSTELLO** is an instructor of education at East Carolina University. She has been a teacher, district English language arts/reading coordinator, and leader for a teacher-research-based curriculum and assessment project. Her teaching and research focus on developing critical literacy practices and investigating the consequences of these practices for individuals. She can be reached at Department of Curriculum and Instruction, East Carolina University, Speight Building, Greenville, NC 27858, USA, or by e-mail at costello@mail.ecu.edu.

## REFERENCES

ALLINGTON, R.L. (Ed.). (2002). *Big brother and the national reading curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.

- ALLINGTON, R.L., & JOHNSTON, P.H. (Eds.). (2002). *Reading to learn: Lessons from exemplary fourth-grade classrooms*. New York: Guilford.
- ALLINGTON, R.L., & MCGILL-FRANZEN, A. (1989). Different programs, indifferent instruction. In A. Gartner & D. Lipsky (Eds.), *Beyond separate education* (pp. 75–98). Baltimore: Brookes.
- ALLINGTON, R.L., & MCGILL-FRANZEN, A. (1995). Flunking: Throwing good money after the bad. In R.L. Allington & S.A. Walmsley (Eds.), *No quick fix: Rethinking literacy programs in America's elementary schools* (pp. 45–60). New York: Teachers College Press.
- ALLINGTON, R.L., & WOODSIDE-JIRON, H. (1999). The politics of literacy teaching: How "research" shaped educational policy. *Educational Researcher*, 28(8), 4–13.
- ARGYRIS, C. (1990). *Overcoming organizational defenses: Facilitating organizational learning*. Boston: Allyn & Bacon.
- ARYA, P. (2003). Influences of reading group experiences on second graders' perceptions of themselves as readers. *Literacy Teaching and Learning*, 8(1), 1–18.
- BARRS, M., ELLIS, S., HESTER, H., & THOMAS, A. (1989). *The primary language record: Handbook for teachers*. London: Inner London Education Authority/Centre for Language in Primary Education.
- BAUDANZA, L. (2001). *Disabilities of a child or disabilities of the system?* Unpublished manuscript, University at Albany, Albany, NY.
- BLACK, P., & WILLIAM, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- BLACK, P., & WILLIAM, D. (1998b, October). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa International*, pp. 139–148.
- BOURDIEU, P. (1991). *Language and symbolic power* (J.B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Cambridge, MA: Harvard University Press.
- BRANDT, D. (2001). *Literacy in American lives*. Cambridge, UK: Cambridge University Press.
- BROADFOOT, P. (2002). Editorial. Assessment for lifelong learning: Challenges and choices. *Assessment in Education*, 9(1), 5–7.
- BROIKOU, K. (1992). *Understanding primary grade classroom teachers' special education referral practices*. Unpublished doctoral dissertation, State University of New York at Albany, Albany, NY.
- CARR, M., & CLAXTON, G. (2002). Tracking the development of learning dispositions. *Assessment in Education*, 9(1), 9–37.
- CLAXTON, G. (1999). *Wise up: The challenge of lifelong learning*. New York: Bloomsbury.
- CLAY, M.M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies*, 22, 155–173.
- CLAY, M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- CLAY, M.M. (1993). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.
- COLLINS, J., & BLOT, R.K. (2003). *Literacy and literacies: Texts, power, and identity*. New York: Cambridge University Press.
- COMEYRAS, M. (1995). What can we learn from students' questions? *Theory Into Practice*, 34, 101–106.
- CROOKS, T. (2001, September). *The validity of formative assessments*. Paper presented at the annual meeting of the British Educational Research Association, Leeds, UK.
- CROOKS, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.
- DARLING-HAMMOND, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047–1085.
- DECI, E.L., SIEGEL, N.H., RYAN, R.M., KOESTNER, R., & KAUFFMAN, M. (1982). Effects of performance standards on teaching styles: Behavior of controlling teachers. *Journal of Educational Psychology*, 74, 852–859.
- DELANDSHERE, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education*, 8, 113–133.
- DWECK, C.S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.
- FAIRCLOUGH, N. (1992). *Discourse and social change*. London: Longman.
- FALK, B. (1998). Using direct evidence to assess student progress: How the Primary Language Record supports teaching and learning. In C. Harrison & T. Salinger (Eds.), *Assessing reading 1: Theory and practice: International perspectives on reading assessment* (pp. 152–165). London: Routledge.
- FALK, B. (2001). Professional learning through assessment. In A. Lieberman & L. Miller (Eds.), *Teachers caught in the action: Professional development that matters* (pp. 118–140). New York: Teachers College Press.
- FALK, B., & DARLING-HAMMOND, L. (1993). *The Primary Language Record at P.S. 261: How assessment transforms teaching and learning*. New York: The National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.
- FENNIMORE, B.S. (2000). *Talk matters: Refocusing the language of public school*. New York: Teachers College Press.
- FLOCKTON, L., & CROOKS, T. (1996). *National Education Monitoring Project: Reading and speaking: Assessment results: 1996* (No. 6). Dunedin, New Zealand: Educational Assessment Research Unit.
- GEE, J.P. (1996). *Social linguistics and literacies: Ideology in discourses* (2nd ed.). London: Falmer.
- GEE, J.P. (2000). Discourse and sociocultural studies in reading. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 195–207). Mahwah, NJ: Erlbaum.
- GILMORE, A. (2002). Large-scale assessment and teachers' assessment capacity: Learning opportunities for teachers in the National Education Monitoring Project in New Zealand. *Assessment in Education*, 9, 343–361.
- GOODMAN, Y. (1978). Kidwatching: Observing children in the classroom. In A. Jagger & M.T. Smith-Burke (Eds.), *Observing the language learner* (pp. 9–18). Newark, DE: International Reading Association.
- GREENBERG, K.H., & WILLIAMS, L. (2002). Reciprocity and mutuality in dynamic assessment: Asking uncomfortable questions. In G.M. v. d. Aalsvoort, W.C.M. Resing, & A.J.J.M. Ruijsenaars (Eds.), *Learning potential assessment and cognitive training* (Vol. 7, pp. 91–110). Amsterdam: JAI.
- GREENE, M. (1985). The role of education in democracy. *Educational Horizons*, 63, 3–9.
- HARLEN, W., & CRICK, R.D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169–207.
- HILL, C. (2004). Failing to meet the standards: The English language arts test for fourth graders in New York State. *Teachers College Record*, 106, 1086–1123.
- HIMLEY, M., & CARINI, P.F. (Eds.). (2000). *From another angle: Children's strengths and school standards: The Prospect Center's descriptive review of the child*. New York: Teachers College Press.
- INTERNATIONAL READING ASSOCIATION AND NATIONAL COUNCIL OF TEACHERS OF ENGLISH JOINT TASK FORCE ON ASSESSMENT. (1994). *Standards for the assessment of reading and writing*. Newark, DE: International Reading Association.
- JOHNSON, D., & KRESS, G. (2003). Globalisation, literacy and society: Redesigning pedagogy and assessment. *Assessment in Education*, 10(1), 5–14.
- JOHNSTON, P.H. (1989). Constructive evaluation and the improvement of teaching and learning. *Teachers College Record*, 90, 509–528.
- JOHNSTON, P.H. (1993). Assessment as social practice. In D. Leu & C. Kinzer (Eds.), *42nd yearbook of the National Reading Conference* (pp. 11–23). Chicago: National Reading Conference.
- JOHNSTON, P.H. (1997). *Knowing literacy: Constructive literacy assessment*. York, ME: Stenhouse.
- JOHNSTON, P.H. (1999). Unpacking literate achievement. In J. Gaffney & B. Askew (Eds.), *Stirring the waters: A tribute to Marie Clay* (pp. 17–25). Portsmouth, NH: Heinemann.
- JOHNSTON, P.H. (2003). Assessment conversations. *The Reading Teacher*, 57, 90–92.
- JOHNSTON, P.H. (2004). *Choice words: How our language affects children's learning*. York, ME: Stenhouse.
- JOHNSTON, P.H., AFFLERBACH, P., & WEISS, P. (1993). Teachers' evaluation of teaching and learning of literacy. *Educational Assessment*, 1(2), 91–117.
- JOHNSTON, P.H., JIRON, H.W., & DAY, J.P. (2001). Teaching and learning literate epistemologies. *Journal of Educational Psychology*, 93(1), 223–233.
- JOHNSTON, P.H., & ROGERS, R. (2001). Early literacy assessment. In S.B. Neuman & D.K. Dickenson (Eds.), *Handbook of early literacy research* (pp. 377–389). New York: Guilford.
- KALANTZIS, M., COPE, B., & HARVEY, A. (2003). Assessing multiliteracies and the new basics. *Assessment in Education: Principles, Policy & Practice*, 10, 15–26.
- LEARNING, S. (2002). *FastForWord*. Retrieved August 3, 2002, from <http://www.scilearn.com>.
- MCDERMOTT, R.P. (1993). The acquisition of a child by a learning disability. In S. Chaiklin & J. Lave (Eds.), *Understanding practice:*

- Perspectives on activity and context* (pp. 269–305). Cambridge, UK: Cambridge University Press.
- MCDONALD, B., & BOUD, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in education: Principles, Policy & Practice*, 10(2), 209–220.
- MCNEIL, L.M. (2000). *Contradictions of school reform: Education costs of standardized tests*. New York: Routledge.
- MEHAN, H. (1993). Beneath the skin and between the ears: A case study in the politics of representation. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 241–268). Cambridge, UK: Cambridge University Press.
- MESSICK, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- MORRISON, K., & JOAN, T.F.H. (2002). Testing to destruction: A problem in a small state. *Assessment in Education*, 9, 289–317.
- MOSS, P., & SCHUTZ, A. (2001). Educational standards, assessment and the search for consensus. *American Educational Research Journal*, 38(1), 37–70.
- MOSS, P.A. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 6–12.
- NIEMI, P., & POSKIPARTA, E. (2002). Shadows over phonological awareness training: Resistant learners and dissipating gains. In E. Hjelmqvist & C.V. Euler (Eds.), *Dyslexia and literacy* (pp. 84–99). London, UK: Whurr.
- NO CHILD LEFT BEHIND ACT OF 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- NYSTRAND, M., GAMORAN, A., KACHUR, R., & PRENDERGAST, C. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.
- PARIS, S.G. (2002). Measuring children's reading development using leveled texts. *The Reading Teacher*, 55, 168–170.
- PRADL, G.M. (1996). Reading and democracy: The enduring influence of Louise Rosenblatt. *The New Advocate*, 9(1), 9–22.
- PRESSLEY, M., ALLINGTON, R.L., WHARTON-MACDONALD, R., COLLINS-BLOCK, C., & MORROW, L. (2001). *Learning to read: Lessons from exemplary first-grade classrooms*. New York: Guilford.
- REX, L.A., & NELSON, M.C. (2004). How teachers' professional identities position high-stakes test preparation in their classrooms. *Teachers College Record*, 106, 1288–1331.
- RODGERS, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842–866.
- ROGERS, R. (2002). Between contexts: A critical analysis of family literacy, discursive practices, and literate subjectivities. *Reading Research Quarterly*, 37, 248–277.
- ROGERS, R. (2003). *A critical discourse analysis of family literacy practices: Power in and out of print*. Mahwah, NJ: Erlbaum.
- ROSKOS, K., & NEUMAN, S.B. (1993). Descriptive observations of adults' facilitation of literacy in young children's play. *Early Childhood Research Quarterly*, 8, 77–97.
- RUEDA, R., & MERCER, J. (1985, June). *Predictive analysis of decision-making practices with limited English proficient handicapped students*. Paper presented at the Third Annual Symposium: Exceptional Hispanic Children and Youth. Monograph series, Denver, CO.
- RUSSELL, M., & ABRAMS, L. (2004). Instructional uses of computers for writing: The effect of state testing programs. *Teachers College Record*, 106, 1332–1357.
- SADLER, R.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13, 191–209.
- SCANLON, D.M., VELLUTINO, F.R., SMALL, S.G., & FANUELE, D.P. (2000, April). *Severe reading difficulties—can they be prevented? A comparison of prevention and intervention approaches*. Paper presented at the American Educational Research Association, New Orleans, LA.
- SCHAFFER, H.R. (1996). Joint involvement episodes as context for development. In H. Daniels (Ed.), *An introduction to Vygotsky* (pp. 251–280). London: Routledge.
- SHAVELSON, R.J., BAXTER, G.P., & PINE, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*, 21(4), 22–27.
- SHEPARD, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20(7), 2–16.
- SHEPARD, L.A., TAYLOR, G.A., & KAGAN, S.L. (1996). *Trends in early childhood assessment policies and practices*. Washington, DC: Office of Educational Research & Improvement.
- SMITH, M.L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8–11.
- SMITH, M.L., & ROTTENBERG, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10(4), 7–11.
- SMITH, P. (1997). *A third chance to learn: The development and evaluation of specialized interventions for young children experiencing difficulty with learning to read* (No. 13227). Wellington, New Zealand: National Council for Educational Research.
- STALLMAN, A.C., & PEARSON, P.D. (1991). Formal measures of early literacy. In L.M. Morrow & J.K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 7–44). Englewood Cliffs, NJ: Prentice Hall.
- TEALE, W. (1991). The promise and the challenge of informal assessment in early literacy. In L.M. Morrow & J.K. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 45–61). Englewood Cliffs, NJ: Prentice Hall.
- TORGESON, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, 15(1), 55–64.
- TUDGE, J., & ROGOFF, B. (1989). Peer influences on cognitive development: Piagetian-Vygotskian perspectives. In M.H. Bornstein & J.S. Bruner (Eds.), *Interactions in human development* (pp. 17–40). Hillsdale, NJ: Erlbaum.
- WIXSON, K.K., & PEARSON, P.D. (1998). Policy and assessment strategies to support literacy instruction for a new century. *Peabody Journal of Education*, 74, 202–227.
- YALOM, I.D. (1989). *Love's executioner and other tales of psychotherapy*. New York: Basic Books.