
The Role of Assessment in a Learning Culture

LORRIE A. SHEPARD

Educational Researcher, Vol. 29, No. 7, pp. 4–14

This article is about classroom assessment—not the kind of assessments used to give grades or to satisfy the accountability demands of an external authority, but rather the kind of assessment that can be used as a part of instruction to support and enhance learning. On this topic, I am especially interested in engaging the very large number of educational researchers who participate, in one way or another, in teacher education. The transformation of assessment practices cannot be accomplished in separate tests and measurement courses, but rather should be a central concern in teaching methods courses.

The article is organized in three parts. I present, first, an historical framework highlighting the key tenets of social efficiency curricula, behaviorist learning theories, and “scientific measurement.” Next, I offer a contrasting social-constructivist conceptual framework that blends key ideas from cognitive, constructivist, and sociocultural theories. In the third part, I elaborate on the ways that assessment practices should change to be consistent with and support social-constructivist pedagogy.

The impetus for my development of an historical framework was the observation by Beth Graue (1993) that “assessment and instruction are often conceived as *curiously separate* in both time and purpose” (p. 291, emphasis added). As Graue notes, the measurement approach to classroom assessment, “exemplified by standardized tests and teacher-made emulations of those tests,” presents a barrier to the implementation of more constructivist approaches to instruction.

To understand the origins of Graue’s picture of separation and to help explain its continuing power over present-day practice, I drew the chronology in Figure 1. A longer-term span of history helps us see that those measurement perspectives, now felt to be incompatible with instruction, came from an earlier, highly consistent theoretical framework (on the left) in which conceptions of “scientific measurement” were closely aligned with traditional curricula and beliefs about learning. To the right is an emergent, constructivist paradigm in which teachers’ close assessment of students’ understandings, feedback from peers, and student self-assessments would be a central part of the social processes that mediate the development of intellectual abilities, construction of knowledge, and formation of students’ identities. The best way to understand dissonant current practices, shown in the middle of the figure, is to realize that instruction (at least in its ideal form) is drawn from the emergent paradigm, while testing is held over from the past.

Historical Perspectives: Curriculum, Psychology, and Measurement

The historical framework I present here is familiar to you. Yet, it is important to remind ourselves where traditional views of testing came from and to appreciate how tightly entwined these views of testing are with past models of curriculum and instruction—because dominant theories of the past continue to operate as the default framework affecting and driving current practices and perspectives. Belief systems of teachers, parents, and policymakers derive from these old theories.

A more elaborated version of the paradigm that has predominated throughout the 20th century can be shown as a set of interlocking circles (Figure 2). The central ideas of social efficiency and scientific management in the curriculum circle were closely linked, respectively, to hereditarian theories of individual differences and to associationist and behaviorist learning theories. These psychological theories were, in turn, served by scientific measurement of ability and achievement.

In the early 1900s, the social efficiency movement grew out of the belief that science could be used to solve the problems of industrialization and urbanization. According to social efficiency theory, modern principles of scientific management, intended to maximize the efficiency of factories, could be applied with equal success to schools. This meant taking F. W. Taylor’s example of a detailed analysis of the movements performed by expert bricklayers and applying similar analyses to every vocation for which students were being prepared (Kleibard, 1995). Then, given the new associationist or connectionist psychology with its emphasis on fundamental building blocks, every step would have to be taught specifically. Precise standards of measurement were required to ensure that each skill was mastered at the desired level. And because it was not possible to teach every student the skills of every vocation, scientific measures of ability were also needed to predict one’s future role in life and thereby determine who was best suited for each endeavor. For John Franklin Bobbitt, a leader in the social efficiency movement, a primary goal of curriculum design was the elimination of waste (1912), and it was wasteful to teach people things they would never use. Bobbitt’s most telling principle was that each individual should be educated

LORRIE A. SHEPARD is a professor in the School of Education, Campus Box 249, University of Colorado at Boulder, Boulder, CO 80309-0249. Her area of specialization is psychometrics and the uses and misuses of tests in educational settings. This article was presented as the presidential address at the 2000 AERA Annual Meeting in New Orleans, LA.

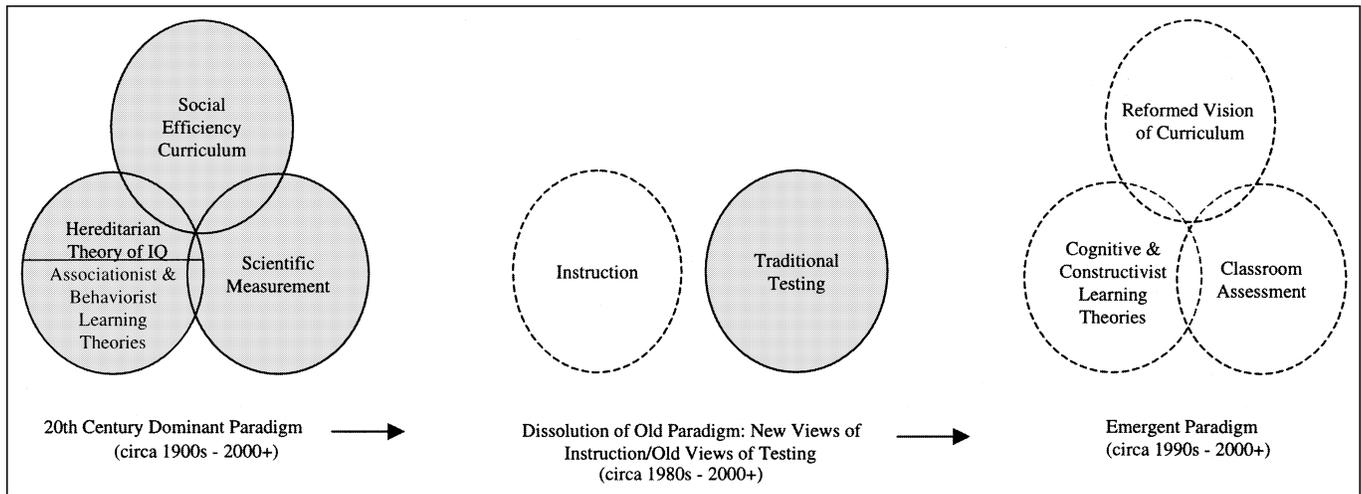


FIGURE 1. An historical overview illustrating how changing conceptions of curriculum, learning theory, and measurement explain the current incompatibility between new views of instruction and traditional views of testing.

“according to his capabilities.” These views led to a highly differentiated curriculum and a largely utilitarian one that disdained academic subjects for any but college preparatory students.

Alongside these curriculum theories, Edward Thorndike’s (1922) associationism and the behaviorism of Hull (1943), Skinner (1938, 1954) and Gagne (1965) conceived of learning as the accumulation of stimulus-response associations. The following quotation from B. F. Skinner is illustrative:

The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength. . . . By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum. (Skinner, 1954, p. 94)

Note that this viewpoint promotes a theory of motivation as well as one of cognitive development.

Several key assumptions of the behavioristic model had consequences for ensuing conceptualizations of teaching and testing:

1. Learning occurs by accumulating atomized bits of knowledge;
2. Learning is tightly sequenced and hierarchical;
3. Transfer is limited, so each objective must be explicitly taught;
4. Tests should be used frequently to ensure mastery before proceeding to the next objective;
5. Tests are isomorphic with learning (tests = learning);
6. Motivation is external and based on positive reinforcement of many small steps.

It is no coincidence that Thorndike was both the originator of associationist learning theory and the “father” of “scientific measurement,” a name given him by Ayers in 1918. Thorndike and his students fostered the development and dominance of the “objective” test, which has been the single most striking feature of achievement testing in the United

States from the beginning of the century to the present day. Recognizing the common paternity of behaviorist learning theory and objective testing helps us to understand the continued intellectual kinship between one-skill-at-a-time test items and instructional practices aimed at mastery of constituent elements.

Looking at any collection of tests from early in the century, as shown in Figure 3, one is immediately struck by how much the questions emphasized rote recall. To be fair, at the time, this was not a distortion of subject matter caused by the adoption of objective-item formats. One hundred years ago, various recall, completion, matching, and multiple-choice test types, along with some essay questions, fit closely with what was deemed important to learn. However, once curriculum became encapsulated and represented by these types of items, it is reasonable to say that these formats locked in a particular and outdated conception of subject matter.

The dominance of objective tests in classroom practice has affected more than the form of subject-matter knowledge. It has also shaped beliefs about the nature of evidence and principles of fairness. In a recent assessment project, for example, both teachers and researchers were surprised to find that despite our shared enthusiasm for developing alternatives to standardized tests we nonetheless operated from different assumptions about how “standardized” assessments needed to be in classrooms. More surprising still, it was teachers who held beliefs more consistent with traditional principles of scientific measurement. From the perspective of our teacher colleagues, assessment needed to be an official event, separate from instruction (Bliem & Davinroy, 1997). To ensure fairness, teachers believed that assessments had to be *uniformly* administered, so they were reluctant to conduct more intensive individualized assessments with only below-grade-level readers. Because of the belief that assessments had to be targeted to a specific instructional goal, teachers felt more comfortable using two separate assessments for separate goals, “running records” to assess fluency and written summaries to assess comprehension rather than, say, asking students to retell the gist of a story in conjunction with running records. Most signifi-

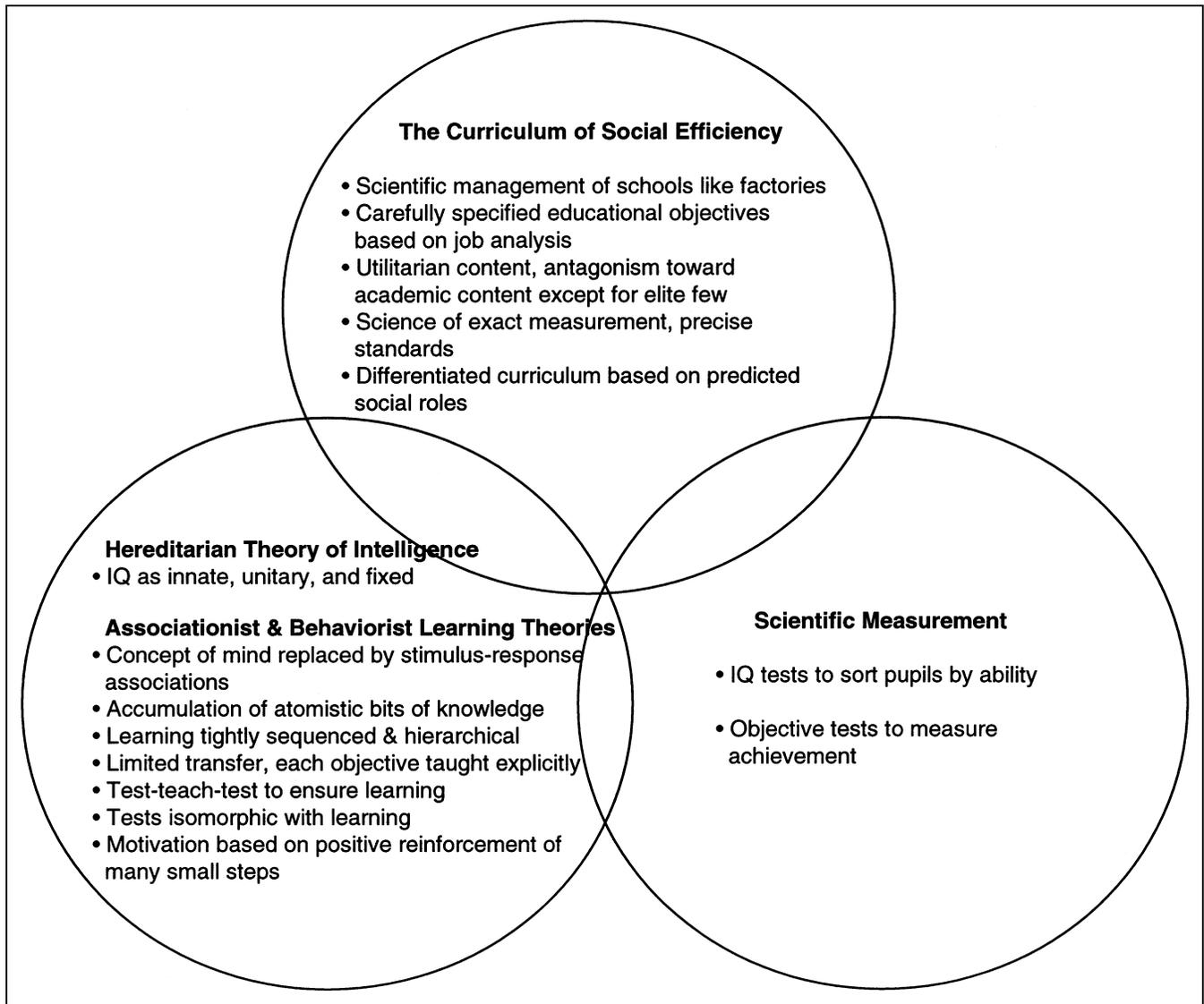


FIGURE 2. *Interlocking tenets of curriculum theory, psychological theories, and measurement theory characterizing the dominant 20th-century paradigm.*

cantly, teachers wanted their assessments to be “objective,” and this was the word they used. They worried often about the subjectivity involved in making more holistic evaluations of student work and preferred formula-based methods, such as counting miscues, because these techniques were more “impartial.”

Any attempt to change the form and purpose of classroom assessment to make it more fundamentally a part of the learning process must acknowledge the power of these enduring and hidden beliefs.

Conceptual Framework: New Theories of Curriculum, Learning, and Assessment

To consider how classroom assessment practices might be reconceptualized to be more effective in moving forward the teaching and learning process, I elaborated the principles of a “social-constructivist” conceptual framework, borrowing from cognitive, constructivist, and sociocultural theories.¹ (Though these camps are sometimes warring with each

other, I predict that it will be something like this merged, middle-ground theory that will eventually be accepted as common wisdom and carried into practice.) The three-part figure (Figure 4) was developed in parallel to the three-part historical paradigm to highlight, respectively, changes in curriculum, learning theory, and assessment. In some cases, principles in the new paradigm are the direct antitheses of principles in the old. The interlocking circles again are intended to show the coherence and inter-relatedness of these ideas taken together.

The cognitive revolution reintroduced the concept of mind. In contrast to past, mechanistic theories of knowledge acquisition, we now understand that learning is an active process of mental construction and sense making. From cognitive theory we have also learned that existing knowledge structures and beliefs work to enable or impede new learning, that intelligent thought involves self-monitoring and awareness about when and how to use skills, and that “expertise” develops in a field of study as a principled and

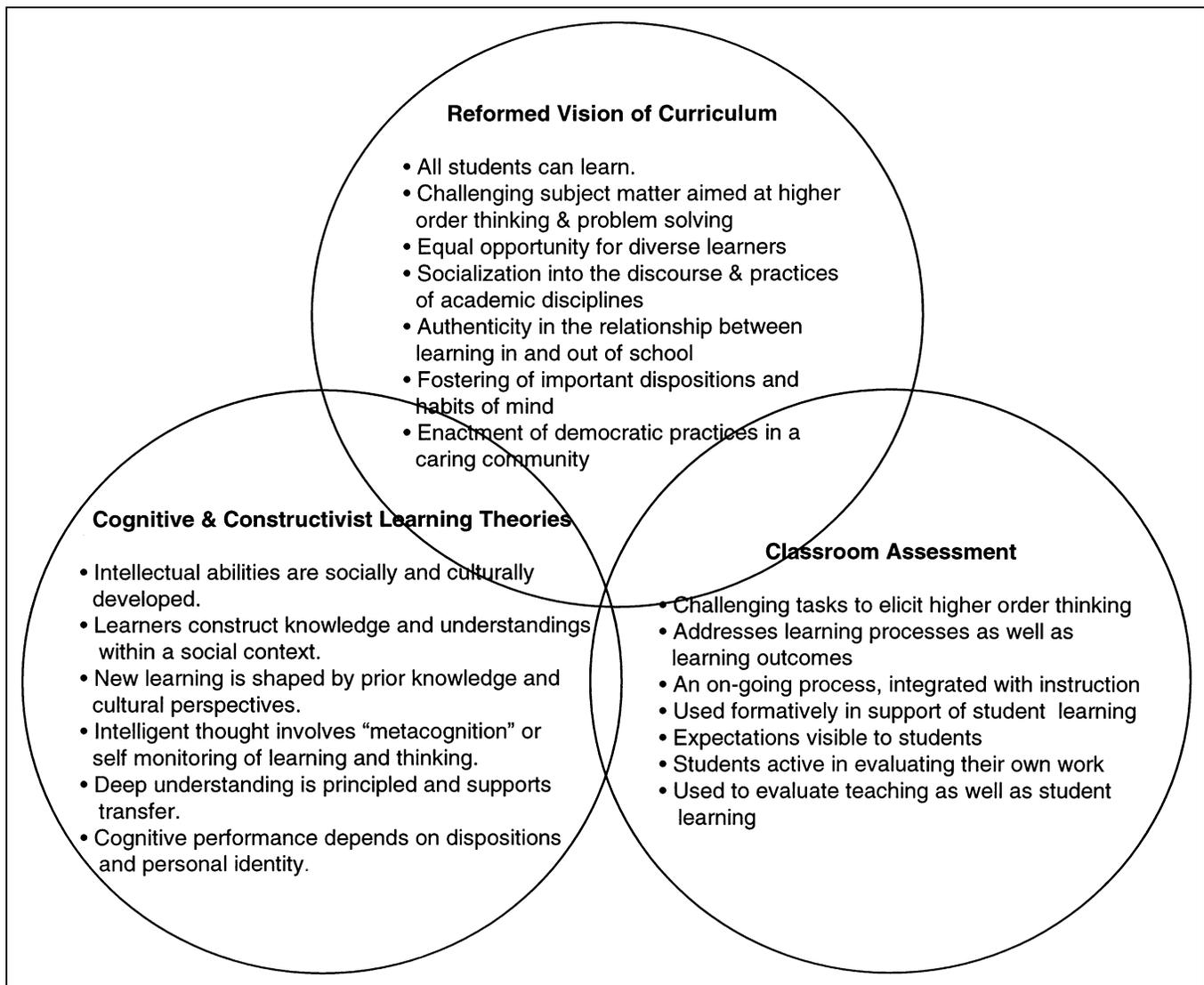


FIGURE 4. Shared principles of curriculum theories, psychological theories and assessment theory characterizing an emergent, constructivist paradigm.

means to know and learn in each of the disciplines. Therefore, a broader range of assessment tools is needed to capture important learning goals and processes and to more directly connect assessment to ongoing instruction. The most obvious reform has been to devise more open-ended performance tasks to ensure that students are able to reason critically, to solve complex problems, and to apply their knowledge in real-world contexts. In addition, if instructional goals include developing students’ metacognitive abilities, fostering important dispositions, and socializing students into the discourse and practices of academic disciplines, then it is essential that classroom routines and corresponding assessments reflect these goals as well. This means expanding the armamentarium for data gathering to include observations, clinical interviews, reflective journals, projects, demonstrations, collections of student work, and students’ self-evaluations, and it means that teachers must engage in systematic analysis of the available evidence.

In this article, I do not elaborate further on needed changes in the content and form of assessment primarily be-

cause this aspect of reform has received the most attention to date. Although I cannot claim that common practice has moved significantly beyond the end-of-chapter test, there are nonetheless already promising models being developed and used in literacy, mathematics, science, history, and so forth. For example, Pat Thompson (1995) provided the set of questions in Figure 5 to illustrate how non-algorithmic problems can help students “see” a mathematical idea. Two additional open-ended tasks are shown in Figure 6 and serve to illustrate the point that good assessment tasks are interchangeable with good instructional tasks.

Protecting Classroom Assessment From the Negative Effects of High-Stakes Accountability Testing

The arguments advanced thus far—in support of social-constructivist learning theory, challenging curriculum for all students, and imaginative new forms of assessment—follow closely the rhetoric of standards-based reform. I have avoided using that term, however, because, from the beginning, standards-based reform has additionally placed

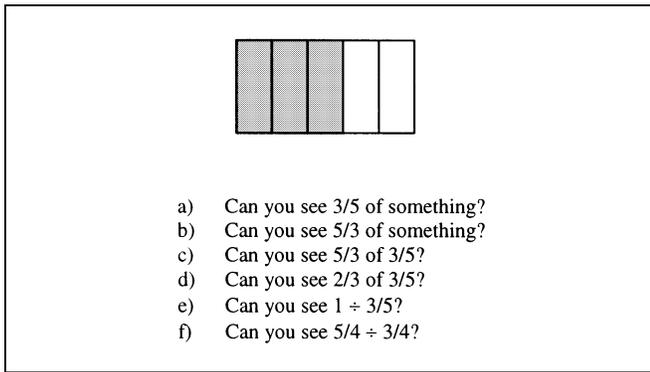


FIGURE 5. An example of a set of questions designed to help students visualize part-whole relationships as a way to understand fractions (Thompson, 1995).

great faith in externally imposed standards and “tests worth teaching to.” More recently, the standards movement has been corrupted, in many instances, into a heavy-handed system of rewards and punishments without the capacity building and professional development originally proposed as part of the vision (McLaughlin & Shepard, 1995). Although both large-scale, system-monitoring assessments and classroom assessments could benefit from the same kinds of substantive reform and alignment of content with important learning goals, there is more at stake here than reform of assessment format. If we wish to pursue seriously

the use of assessment *for* learning, which I consider in the next section, it is important to recognize the pervasive negative effects of accountability tests and the extent to which externally imposed testing programs prevent and drive out thoughtful classroom practices. In presenting these ideas to an audience of educational researchers and teacher educators, I used the image of Darth Vader and the Death Star to convey the overshadowing effects of accountability testing.

The negative effects of high-stakes testing on teaching and learning are well known (e.g., Madaus, West, Harmon, Lomax, & Viator, 1992). Under intense political pressure, test scores are likely to go up without a corresponding improvement in student learning. In fact, distortions in what and how students are taught may actually decrease students’ conceptual understanding. While some had imagined that teaching to good tests would be an improvement over low-level basic-skills curricula, more recent experiences remind us that all tests can be corrupted. And all can have a corrupting influence on teaching (Whitford & Jones, 2000).

Moreover, as Darling-Hammond (1988), McNeil (1988), and others have pointed out, external accountability testing leads to the de-skilling and de-professionalization of teachers, even—in my own state recently—to the denigration of teaching. High-stakes accountability teaches students that effort in school should be in response to externally administered rewards and punishment rather than the excitement of ideas. And accountability-testing mandates warn teachers to comply or get out (or move, if they can, to schools with higher scoring students).

Grade 4 Mathematics Problem Set (Mathematical Sciences Education Board, 1993)	Grade 5 Science Tasks (California Learning Assessment System, 1994)
<p>All of the bridges in this part are built with yellow rods for spans and red rods for supports, like the one shown here. This is a 2-span bridge like the one you just built. Note that the yellow rods are 5 cm long.</p> <div style="text-align: center; margin: 10px 0;"> </div> <ol style="list-style-type: none"> 1. Now, build a 3-span bridge. <ol style="list-style-type: none"> a. How many yellow rods did you use? ____ b. How long is your bridge? ____ c. How many red rods did you use? ____ d. How many rods did you use altogether? ____ 2. Try to answer these questions without building a 5-span bridge. If you want, build a 5-span bridge to check your answers. <ol style="list-style-type: none"> a. How many yellow rods would you need for a 5-span bridge? ____ b. How long would your bridge be? ____ c. How many red rods would you need? ____ d. How many rods would you need altogether? ____ 3. Write a rule for figuring out the total number of rods you would need to build a bridge if you knew how many spans the bridge had. 	<p>Fossils You are a paleontologist (a scientist who studies past life forms). You were digging and just discovered a large group of fossils.</p> <p>Directions: Open BAG A and spread the fossils on the table. Use the hand lens to carefully observe each fossil. Sort your fossils into groups. You may make as many groups as you like.</p> <p>Write answers to these questions in your journal.</p> <ol style="list-style-type: none"> 1. Draw your groups. Circle and number each group. 2. How many groups do you have? 3. List the number of each group and tell why you sorted your fossils into these groups. <p>=====</p> <p>BAG B has a fossil that was found in the area near where you were digging.</p> <p>Directions: Open BAG B. Take out the new fossil and compare it with the other fossils on the table.</p> <ol style="list-style-type: none"> 4. Does this new fossil fit into one of your groups? If YES, how are they alike? 5. If the new fossil does not fit into any of your groups, describe a new group in which this fossil would fit. 6. Choose one of the fossils and draw a picture of it. 7. In what kind of habitat (environment) do you think this fossil might have once lived? Why?

FIGURE 6. Examples of open-ended assessment tasks intended to engage students in thinking and reasoning about important content.

Again, these ideas are not new. It is likely that teacher educators say something about this litany of complaints in teacher preparation courses. But, what do diatribes against testing teach candidates about more meaningful forms of assessment? Given their own personal histories, our students are able to hate standardized testing and at the same time reproduce it faithfully in their own pre-post testing routines, if they are not given the opportunity to develop and try out other meaningful forms of assessment situated in practice. So we must teach them how to do assessment well.

Also, teacher candidates need to find support and a way of protecting their own developing understandings of constructivist assessment practices from the onslaught of test-driven curricula. I have in mind here something like the double-entry teaching that teachers had invented in Linda McNeil's (1988) study of the *Contradictions of Control*. In contrast to teachers who trivialized content and taught defensively as a means to control and win compliance from students, McNeil found that excited and engaging teachers in the magnet schools she studied found ways to resist and hold off the pernicious effects of proficiency testing on their curriculum. Specifically, they helped students keep parallel sets of notes, one set for the real knowledge and one for the knowledge they would need for the test. They did this rather than give over the entire course to the "fragments and facts" required on the test.

This is only one example of a strategy for resistance. As I continue next to describe productive ways to use assessment in classrooms, I emphasize the need sometimes to "mark" informal assessment occasions for students as they occur within the normal flow of classroom discourse—because this helps students become self-aware about how assessment can help learning. Similarly, I believe we should explicitly address with our teacher education students how they might cope with the contesting forces of good and evil assessment as they compete in classrooms to control curriculum, time, and student attitudes about learning.

Using Assessment in the Process of Learning

A Learning Culture

Improving the content of assessments is important but not sufficient to ensure that assessment will be used to enhance learning. In this section, I consider the changes in classroom practices that are also needed to make it possible for assessment to be used as part of the learning process. How might the culture of classrooms be shifted so that students no longer feign competence or work to perform well on the test as an end separate from real learning? Could we create a learning culture where students and teachers would have a shared expectation that finding out what makes sense and what doesn't is a joint and worthwhile project, essential to taking the next steps in learning?

I believe that our international colleagues are ahead of us in thinking about the difficulties of making these cultural changes. Sadler (1998) in Australia, for example, writes about "the long-term exposure of students to defective patterns of formative² assessment" (p. 77). Perrenoud in Switzerland (1991) notes that there are always certain students in a class who are willing to work harder to learn more and who, therefore, go along with formative assessment. But other children and adolescents are "imprisoned in the identity of a bad pupil and an opponent" (p. 92). According to Perrenoud,

"every teacher who wants to practice formative assessment must reconstruct the teaching contract so as to counteract the habits acquired by his pupils" (p. 92). Tunstall and Gipps (1996) have studied classrooms in Great Britain where teachers have developed more interactive ways of discussing work and criteria with students as a means to redistribute power and establish more collaborative relationships with students.

To accomplish the kind of transformation envisioned, we have not only to make assessment more informative, more insightfully tied to learning steps, but at the same time we must change the social meaning of evaluation. Our aim should be to change our cultural practices so that students and teachers look to assessment as a source of insight and help instead of an occasion for meting out rewards and punishments. In the paragraphs that follow, I summarize briefly several specific assessment strategies: dynamic assessment, assessment of prior knowledge, the use of feedback, teaching for transfer, explicit criteria, student self-assessment, and evaluation of teaching. Each of these strategies serves a social, motivational purpose as well as a cognitive, informational one. None of these strategies by themselves will be effective if they are not part of a more fundamental shift in classroom practices and expectations about learning.

Dynamic, On-Going Assessment

In order for assessment to play a more useful role in helping students learn it should be moved into the middle of the teaching and learning process instead of being postponed as only the end-point of instruction. Dynamic assessment—finding out what a student is able to do independently as well as what can be done with adult guidance—is integral to Vygotsky's idea of a zone of proximal development. This type of interactive assessment, which allows teachers to provide assistance as part of assessment, does more than help teachers gain valuable insights about how understanding might be extended. It also creates perfectly targeted occasions to teach and provides the means to scaffold next steps. Although formal dynamic assessments are assumed to involve an adult working with only one child, these ideas about social mediation of learning can be extended to groups, especially if students are socialized into the ways of talking in a community of practice and become accustomed to explaining their reasoning and offering and receiving feedback about their developing competence as part of a social group.

Note that these ideas, based on activity theory and Lave and Wenger's (1991) concept of legitimate peripheral participation, provide a profoundly different view of motivation from behaviorist reinforcement and create no separation between cognitive and motivational goals. According to Lave and Wenger's theory, learning and development of an identity of mastery occur together as a newcomer becomes increasingly adept at participating in a community of practice. If one's identity is tied to group membership, then it is natural to work to become a more competent and full-fledged member of the group.

Prior Knowledge

Prior knowledge and feedback are two well-established ideas, the meaning of which may have to be reexamined as learning theories are changed to take better account of social and cultural contexts. For example, assessing my prior

knowledge using a checklist or pre-test version of the intended end-of-unit test may not be very accurate unless I already have sophisticated experience with the teacher's measures and conceptual categories. Open discussion or "instructional conversations" (Tharp & Gallimore, 1988) are more likely to elicit a more coherent version of students' reasoning and relevant experiences and can be a much more productive way for novice teachers to learn about the resources brought by students from diverse communities.

In my own experience working in schools, I have noticed two divergent sets of teaching practices that address students' prior knowledge. First, many teachers rely on a traditional, pretest-posttest design to document student progress, but then do not use information from the pretest in instruction. At the same time, a significant number of teachers, especially in reading and language arts, use prior knowledge activation techniques, such as Ogle's (1986) KWL strategy, but without necessarily attending to the assessment insights provided.

We have a great deal of work to do to develop and model effective assessment strategies, for starting points as well as for other stages of learning. One question we may want to consider is whether assessment should become so much a part of normal classroom discourse patterns that scaffolding and ongoing checks for understanding are embedded (and therefore disguised). Or whether assessment steps should be marked and made visible to students as an essential step in learning. In our efforts to change the culture of the classroom, it may be helpful, at least in the short term, to label prior knowledge activation techniques as instances of "assessment." What safer time to admit what you don't know than at the start of an instructional activity?

Feedback

We take it for granted that providing feedback to the learner about performance will lead to self-correction and improvement. For the most part, however, the existing literature on feedback will be of limited value to us in reconceptualizing assessment from a constructivist perspective, because the great majority of existing studies are based on behaviorist assumptions. Typically, the outcome measures are narrowly defined, feedback consists of reporting of right and wrong answers to the learner, and the end-of-study test may differ only slightly from the prior measure and from instructional materials.

More promising are studies of scaffolding and naturalistic studies of expert tutoring—but these studies also reveal how much we have to learn about effective use of feedback. For example, Lepper, Drake and O'Donnell-Johnson (1997) found that the most effective tutors do not routinely correct student errors directly. Instead they *ignore* errors when they are inconsequential to the solution process and *forestall* errors that the student has made previously by offering hints or asking leading questions. Only when the forestalling tactic fails do expert tutors *intervene* with a direct question intended to force the student to self-correct, or they may engage in *debugging*, using a series of increasingly direct questions to guide the student through the solution process. According to Lepper et al.'s analysis, the tendency of expert tutors to use indirect forms of feedback when possible was influenced by their desire to maintain student motivation and self-confidence while not ignoring student errors. This is a balancing act that new teachers must learn to perform as well.

Transfer

There is a close relationship between truly *understanding* a concept and being able to *transfer* knowledge and use it in new situations. In contrast to memorization—and in contrast to the behaviorist assumption that each application must be taught as a separate learning objective—true understanding is flexible, connected, and generalizable. Not surprisingly, research studies demonstrate that learning is more likely to transfer if students have the opportunity to practice with a variety of applications while learning (Bransford, 1979). Although there appears to be disagreement between cognitivists and situativists regarding knowledge generalization (Anderson, Reder, & Simon, 1996), in fact, both groups of researchers acknowledge the importance of being able to use what one has learned in new situations (Bransford, Brown, & Cocking, 1999). Cognitivists focus more on cognitive structures, abstract representations, and generalized principles that enable knowledge use in new situations, while situativists are concerned about "learning to participate in interactions in ways that succeed over a broad range of situations" (Greeno, 1996, p. 3).

In working with pre-service teachers, I have suggested that a goal of teaching should be to help students develop "robust" understandings (Shepard, 1997). The term was prompted by Marilyn Burns's (1993) reference to children's understandings as being "fragile"—they appear to know a concept in one context but not to know it when asked in another way or in another setting. Sometimes this fragility occurs because students are still in the process of learning and sometimes because the framing of the problem, clues, and other supports available in the familiar context are not available in another. All too often, however, mastery appears pat and certain but does not travel to new situations because students have mastered classroom routines and not the underlying concepts. To support generalization and ensure transfer, that is, to support robust understandings, "Good teaching constantly asks about old understandings in new ways, calls for new applications, and draws new connections" (Shepard, 1997, p. 27). And good assessment does the same. We should not, for example, agree to a contract with our students which says that the only fair test is one with familiar and well-rehearsed problems.

Explicit Criteria

Frederiksen and Collins (1989) used the term *transparency* to express the idea that students must have a clear understanding of the criteria by which their work will be assessed. In fact, the features of excellent performance should be so transparent that students can learn to evaluate their own work in the same way that their teachers would. According to Frederiksen and Collins,

The assessment system (should) provide a basis for developing a metacognitive awareness of what are important characteristics of good problem solving, good writing, good experimentation, good historical analysis, and so on. Moreover, such an assessment can address not only the product one is trying to achieve, but also the process of achieving it, that is, the habits of mind that contribute to successful writing, painting, and problem solving (Wiggins, 1989). (Frederiksen & Collins, 1989, p. 30)

Having access to evaluation criteria satisfies a basic fairness principle (we should know the rules for how our work

will be judged). More importantly, however, giving students the opportunity to get good at what it is that the standards require speaks to a different and even more fundamental sense of fairness, which is what Wolf and Reardon (1996) had in mind when they talked about “making thinking visible” and “making excellence attainable.”

Self-Assessment

Student self-assessment serves cognitive purposes, then, but it also promises to increase students’ responsibility for their own learning and to make the relationship between teachers and students more collaborative. As Caroline Gipps (1999) has suggested, this does not mean that the teacher gives up responsibility, but that rather, by sharing it, she gains greater student ownership, less distrust, and more appreciation that standards are not capricious or arbitrary. In case studies of student self-evaluation practices in both an Australian and English site, Klenowski (1995) found that students participating in self-evaluation became more interested in the criteria and substantive feedback than in their grade per se. Students also reported that they had to be more honest about their own work as well as being fair with other students, and they had to be prepared to defend their opinions in terms of the evidence. Klenowski’s (1995) data support Wiggins’s (1992) earlier assertion that involving students in analyzing their own work builds ownership of the evaluation process and “makes it possible to hold students to higher standards because the criteria are clear and reasonable” (p. 30).

Evaluation of Teaching

In addition to using assessment to monitor and promote individual students’ learning, classroom assessment should also be used to examine and improve teaching practices. This includes both ongoing, informal assessments of students’ understandings to adjust lessons and teaching plans as well as more formal and critical action-research studies. As I have suggested with other assessment strategies, here again I believe it will be helpful for teachers to make their investigations of teaching visible to students, for example, by discussing with them decisions to redirect instruction, stop for a mini-lesson, and so forth. This seems to be fundamentally important to the idea of transforming the culture of the classroom. If we want to develop a community of learners—where students naturally seek feedback and critique their own work—then it is reasonable that teachers would model this same commitment to using data systematically as it applies to their own role in the teaching and learning process.

Conclusion

In conclusion, let me acknowledge that this social-constructivist view of classroom assessment is an idealization. The new ideas and perspectives underlying it have a basis in theory and empirical studies, but how they will work in practice and on a larger scale is not known. Clearly, the abilities needed to implement a reformed vision of curriculum and classroom assessment are daunting. Being able to ask the right questions at the right time, anticipate conceptual pitfalls, and have at the ready a repertoire of tasks that will help students take the next steps requires deep knowledge of subject matter. Teachers will also need help in learning to use assessment in new ways. They will need a theory of mo-

tivation and a sense of how to develop a classroom culture with learning at its center. Given that new ideas about the role of assessment are likely to be at odds with prevailing beliefs, teachers will need assistance to reflect on their own beliefs as well as those of students, colleagues, parents, and school administrators.

I am reminded of Linda Darling-Hammond’s (1996) acknowledgement in her presidential address that John Dewey anticipated all of these ideas 100 years ago. But as Cremin (1961) explained, the successes of progressive education reforms never spread widely because such practice required “infinitely skilled teachers” who were never prepared in sufficient numbers to sustain these complex forms of teaching and schooling.

So, we are asking a lot of ourselves and others. Nonetheless, we must try again. This vision should be pursued because it holds the most promise for using assessment to improve teaching and learning. To do otherwise means that day-to-day instructional practices will continue to reinforce and reproduce the status quo. Our goal should be to find ways to fend off the negative effects of externally imposed tests and to develop instead classroom assessment practices that can be trusted to help students take the next steps in learning.

Epilogue

I would be remiss if I did not take this opportunity to provide at least a brief sketch of what we might do concretely to work toward a proposed vision of assessment in the service of learning. Happily for an organization of researchers, I suggest more research—but research of a particular kind embedded in the dilemmas of practice. I also suggest that we develop and pursue an agenda of public education to help policymakers and the general citizenry understand the differences between large-scale, system monitoring tests and what we hope for from teachers on a daily basis.

A Program of Research

To develop effective practices based on social-constructivist perspectives, it will be important to conduct studies in classrooms where instruction and assessment strategies are consonant with this model. In many cases this will mean “starting over again” and not assuming that findings from previous research studies can be generalized across paradigms. For example, as suggested earlier, there are hundreds of studies on feedback but nearly all conform to behaviorist assumptions—instruction is of short duration, posttests closely resemble pretests, feedback is in the form of being told the correct answers, and so forth. New studies will be needed to further our understandings of feedback provided in ways that reflect constructivist principles, for example, as part of instructional scaffolding, assessment conversations, and other interactive means of helping students self-correct and improve. Similarly, the research literature on motivation makes sweeping claims about the risks of evaluating students, especially when they are tackling difficult problems. Yet, these findings are based on students’ experiences with traditional, inauthentic and normative forms of assessment, where students took little responsibility for their own learning, and criteria remained mysterious. If the classroom culture were to be shifted dramatically, consistent with social-constructivist learning perspectives, then the effects of assessing students on difficult problems would

have to be reexamined. Thus we face the challenge of trying to find out what works at the same time that we are attempting to create new contexts and new cultural expectations that will fundamentally alter the very relations we are trying to study.

We also need to study what makes sense in terms of teacher development and change. Many of the most exciting current assessment projects are being conducted in classrooms but still have researchers at the helm, taking central responsibility for the development of curriculum, assessment tasks, and technology-based delivery systems. We know that for teachers to make meaningful changes in pedagogical beliefs and accompanying practices, they themselves will need to try out and reflect on new approaches in the context of their own classrooms (Putnam & Borko, 1997). In deference to the enormous constraints on teachers' time, we should also look for ways to introduce new practices incrementally, for example, to develop a portfolio for one subject area or one curriculum unit before trying to do it in all subject areas. To consider how particular classroom assessment strategies might be used to create a learning culture as well as improve achievement, teams of teachers in schools might undertake projects aimed at any one of the assessment elements. For example, one team might want to introduce self-assessment and conference with students about how (or whether) self-assessment helps them learn. Another team of teachers might agree to meet regularly to share examples of "assessment insights," that is, specific occasions when assessment data from a student, written or oral, helped the teacher intervene in a better way because she understood what the student was thinking. While another group of teachers might focus on using feedback explicitly to help students make their work better.

When I say that our research efforts should be embedded in the dilemmas of practice, I am echoing the call for more collaborative forms of research advanced in recent reports by the National Research Council (1999) and National Academy of Education (1999) as well as by Alan Schoenfeld (1999) in his presidential address to the AERA. In contrast to a traditional, linear progression from research to development and dissemination, these authors argue for investing in research projects that would advance fundamental understandings at the same time that they would work to solve practical problems in real-world settings. If researchers and professional educators share responsibility for improving educational outcomes, it is hoped that research will lead to continuous improvement of practice and not require a separate translation phase to be useful. In the context of an agenda for improving classroom assessment, this model for research would mean conducting studies aimed at general explanatory principles regarding prior knowledge, self-assessment, and the like, at the same time that practical issues are addressed such as the initial obstacles of negative student attitudes, time seemingly stolen from instruction, and the inevitable demand for better materials and instructional tasks that elicit the kind of thinking and dialogue envisioned.

A Public Education Agenda

Researchers in the United States have engaged policymakers and the public on the topic of testing but have focused almost exclusively on the features of state and district accountabil-

ity testing programs—what the content should be, whether there should be high-stakes consequences, and so forth. In contrast, we have much to learn from assessment experts in the United Kingdom who have pursued a fundamentally different course of action emphasizing the key role of formative assessment in effective teaching. Beginning in 1989, researchers representing England, Northern Ireland, Scotland, and Wales met as a Task Group of the British Educational Research Association and ultimately established themselves as the Assessment Reform Group. The group is concerned with policy issues and has attempted to have a dialogue with policymakers. Although members of the group have been involved with either the development or evaluation of the National Assessment Programme, they "have become more and more convinced of the crucial link between assessment, as carried out in the classroom, and learning and teaching" (Assessment Reform Group, 1999, p. 1). They commissioned a major review of research examining the impact of assessment on students' learning (Black & Wiliam, 1998a), and they have issued two policy-oriented "little books" summarizing the important tenets of *assessment for learning* and urging government policies that would give more than lip service to the importance of improving formative assessment (Assessment Reform Group, 1999; Black & Wiliam, 1998b). They have argued for (a) reframing of bureaucratic requirements, such as standards for teacher education and school inspections, to ensure that teachers are skilled assessors of students' learning; (b) increased funding, especially for teacher professional development; and (c) reducing obstacles, especially the influence of external tests that dominate teachers' work.

Assessment experts in the U.S. should consider whether a similar public education endeavor would be worthwhile and what message we would choose to convey. At a minimum, we should try to get beyond the currently popular sound-bite of "instructionally relevant assessment," because, unfortunately, legislators and school board members have taken up this slogan with the intention that once-per-year accountability testing can be used to diagnose individual student needs. Yes, end-of-year tests can be used to evaluate instruction and even tell us something about individual students; but such exams are like shopping mall medical screenings compared to the in-depth and ongoing assessments needed to genuinely increase learning. By pursuing a public education agenda like that undertaken in the U.K. we could help policymakers understand the limits to what can be accomplished with accountability tests (and thereby fend off their negative effects) and at the same time garner the support and flexibility that teachers and researchers will need to develop powerful examples and to enact more pervasive shifts in classroom practices.

Notes

The work reported herein was supported in part by grants from the Office of Educational Research and Improvement, U.S. Department of Education, to the Center for Research on Evaluation, Standards, and Student Testing (CREST) (Award No. R305B60002) and to the Center for Research on Evaluation, Diversity and Excellence (CREDE) (Award No. R306A60001). The findings and opinions expressed in this article do not reflect the positions or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

¹ A more detailed discussion of this framework and supporting literature review are provided in Shepard (in press).

² Sadler (1998) uses the term *formative assessment* to mean assessment “that is specifically intended to provide feedback on performance to improve and accelerate learning” (p. 77). He acknowledges that teachers may have difficulty using feedback in positive ways because of students’ negative coping strategies developed in response to past practices.

References

- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25, 5–11.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge School of Education.
- Au, K. H., & Jordan, C. (1981). Teaching reading to Hawaiian children: Finding a culturally appropriate solution. In H. Trueba, G. P. Guthrie, & K. H. Au (Eds.), *Culture in the bilingual classroom: Studies in classroom ethnography* (pp. 139–152). Rowley, MA: Newbury House.
- Ayers, L. P. (1918). History and present status of educational measurements. *Seventeenth Yearbook of the National Society for the Study of Education, Part II*, 9–15.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London: School of Education, King’s College.
- Bliem, C. L., & Davinroy, K. H. (1997). *Teachers’ beliefs about assessment and instruction in literacy*. Unpublished manuscript, University of Colorado at Boulder.
- Bobbitt, F. (1912). The elimination of waste in education. *The Elementary School Teacher*, 12, 259–271.
- Bransford, J. D. (1979). *Human cognition: Learning, understanding, and remembering*. Belmont, CA: Wadsworth.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Brown, A. L. (1994). The advancement of learning. *Educational Researcher*, 23, 4–12.
- Burns, M. (1993). *Mathematics: Assessing understanding*. White Plains, NY: Cuisenaire Company of America.
- California Learning Assessment System. (1994). *A sampler of science assessment—elementary*. Sacramento: California Department of Education.
- Cremin, L. (1961). *The transformation of the school: Progressivism in American education, 1876–1957*. New York: Vintage Books.
- Darling-Hammond, L. (1988). Accountability and teacher professionalism. *American Educator*, 12, 8–13.
- Darling-Hammond, L. (1996). The right to learn and the advancement of teaching: Research, policy, and practice for democratic education. *Educational Researcher*, 25, 5–17.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18, 27–32.
- Gagne, R. M. (1965). *The conditions of learning*. New York: Rinehard & Winston.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. In P.D. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education* (Vol. 24, pp. 355–392). Washington, DC: American Educational Research Association.
- Graue, M. E. (1993). Integrating theory and practice through instructional assessment. *Educational Assessment*, 1, 293–309.
- Greeno, J. G. (1996, July). *On claims that answer the wrong questions*. Stanford, CA: Institute for Research on Learning.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge: Cambridge University Press.
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century.
- Klenowski, V. (1995). Student self-evaluation process in student-centered teaching and learning contexts of Australia and England. *Assessment in Education*, 2, 145–163.
- Kliebard, H. M. (1995). *The struggle for the American curriculum: 1893–1958* (2nd ed.). New York: Routledge.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Lepper, M. R., Drake, M. F., O’Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches & issues*. Cambridge, MA: Brookline Books.
- Madaus, G. F., West, M. M., Harmon, M. C., Lomax, R. G., & Viator, K. A. (1992). *The influence of testing on teaching math and science in grades 4–12*. Chestnut Hill, MA: Center of Study of Testing, Evaluation, and Educational Policy, Boston College.
- Mathematical Sciences Education Board. (1993). *Measuring up: Prototypes for mathematics assessment*. Washington, DC: National Academy Press.
- McLaughlin, M. W., & Shepard, L. A. (1995). *Improving education through standards-based reform: A report of the National Academy of Education panel on standards-based educational reform*. Stanford, CA: National Academy of Education.
- McNeil, L. M. (1988). *Contradictions of control: School structure and school knowledge*. New York: Routledge.
- National Academy of Education. (1999, March). *Recommendations regarding research priorities: An advisory report to the National Educational Research Policy and Priorities Board*. New York: New York University.
- National Research Council. (1999). *Improving student learning: A strategic plan for education research and its utilization*. Washington, DC: National Academy Press.
- Ogle, D. M. (1986). K-W-L: A teaching model that develops active reading of expository text. *The Reading Teacher*, 39(6), 564–570.
- Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation. In P. Weston (Ed.), *Assessment of pupils’ achievement: Motivation and school success* (pp. 77–101). Amsterdam: Swets and Zeitlinger.
- Putnam, R. T., & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *International handbook of teachers and teaching* (Vol. 2, pp. 1223–1296). Dordrecht, The Netherlands: Kluwer.
- Ross, C. C. (1941). *Measurement in today’s schools*. New York: Prentice-Hall.
- Ruch, G. M. (1929). *The objective or new-type examination*. Chicago: Scott Foresman.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy and Practice*, 5, 77–84.
- Schoenfeld, A. H. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher*, 28(7), 4–14.
- Shepard, L. A. (1997). *Measuring achievement: What does it mean to test for robust understanding?* Princeton, NJ: Policy Information Center, Educational Testing Service.
- Shepard, L. A. (in press). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed). Washington, DC: American Educational Research Association.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24, 86–97.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. New York: Cambridge University Press.
- Thompson, P. W. (1995). Notation, convention, and quantity in elementary mathematics. In J. T. Sowder & B. P. Schappelle (Eds.), *Providing a foundation for teaching mathematics in the middle grades* (pp. 199–221). New York: State University of New York Press.
- Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22, 389–404.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Whitford, B. L., & Jones, K. (2000). Kentucky lesson: How high stakes school accountability undermines a performance-based curriculum vision. In B. L. Whitford & K. Jones (Eds.), *Accountability, assessment, and teacher commitment: Lessons from Kentucky’s reform efforts*. Albany, NY: State University of New York Press.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703–713.
- Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49, 26–33.
- Wolf, D. P., & Reardon, S. F. (1996). Access to excellence through new forms of student assessment. In J. B. Baron & Wolf, D. P. (Eds.), *Performance-based student assessment: Challenges and possibilities* (pp. 1–31). Chicago: University of Chicago Press.