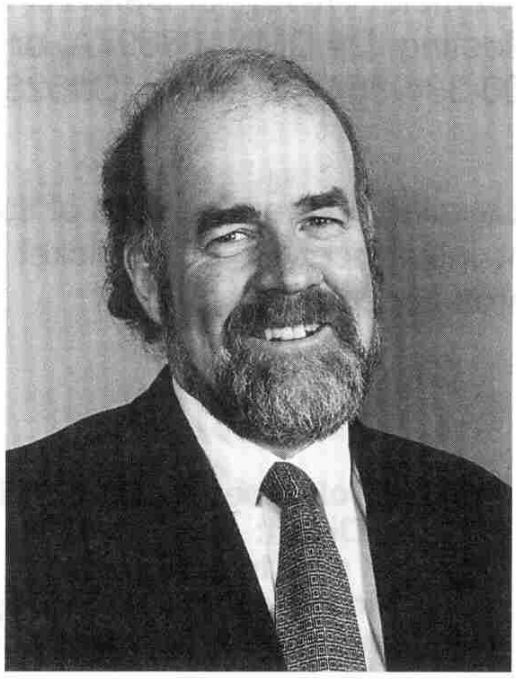
Robert J. Tierney

Literacy assessment reform: Shifting beliefs, principled possibilities, and emerging practices

Distinguished Educator Rob Tierney articulates 13 key principles for literacy assessment.



Developing better assessment practices requires more than simply choosing a new test or adopting a packaged informal assessment procedure. Indeed, it is difficult to imagine "plastic wrapped" versions of what these new assessment systems intend. Unfortunately, some assessment practices may be repackaged versions of old tests rather than new ways of doing assessment. And some assessment practices, regardless of the label (authentic assessment, alternative assessment, student-centered assessment, responsive evaluation, classroom-based assessment, or constructive assessment), may be compromised as they are made to fit tenets or principles out of character or inconsistent with the aspirations of these possibilities. Contributing to the confusion may be reverence for certain technical attributes espoused by some pyschometricians and a predilection or political climate that tends to perpetuate top-down assessment and curriculum reform. Not surprising, professionals may differ in whether or not new forms of assessment live up to their promise.

In hopes of helping to sort out some of these dilemmas—the oxymorons, compromises or, at the very least, different views of assessment, learners and learning, I have tried to make the ramifications of my definition of assessment more explicit with the articulation of a number of principles, which I describe in this article.

(pp. 374–390)

These principles for assessment emanate from personal ideals and practice as much as theory and research—a mix of child-centered views of teaching, pluralistic and developmental views of children, constructivist views of knowing, and critical theoretical views of empowerment. The view that I espouse strives to be in harmony with Bruner's (1990) notion that a democratic society "demands that we be conscious of how we come to our knowledge and be as conscious as we can be about the values that lead us to our perspectives. It asks us to be accountable for how and what we know" (p. 31). Likewise, my goal is aligned with constructivists' ways of knowing and the notion of responsive evaluation that Guba and Lincoln (1989) as well as others (e.g., Lather, 1986; Stake, 1983) have espoused:

> Responsive evaluation is not only responsive for the reason that it seeks out different stakeholder views but also since it responds to those items in the subsequent collection of information. It is quite likely that different stakeholders will hold very different constructions with respect to any particular claim, concern, or issue. As we shall see, one of the major tasks of the evaluator is to conduct the evaluation in such a way that each group must confront and deal with the constructions of all the others, a process we shall refer to as a hermeneutic dialectic. (Guba & Lincoln, 1989, p. 41)

I also find my views aligning with critical theorists (e.g., Baker & Luke, 1991; Freire & Macedo, 1987; Gee, 1990; hooks, 1989, 1994) who suggest that the point of literacy is to reflect upon, and be empowered by, text rather than to be subjugated by it—that literacy contributes to social transformation as we connect with what we read and write, not in acquiescence, but in reaction, reflection, and response.

In accordance with these notions, I contend that to be both accountable and empowered, readers and writers need to be both reflective and pragmatic. To do so, readers and writers need to be inquirers—researching their own selves, considering the consequences of their efforts, and evaluating the implications, worth, and ongoing usefulness of what they are doing or have done. Teachers can facilitate such reflection by encouraging students to keep traces of what they do, by suggesting they pursue ways to depict their journey (e.g., webs or a narrative or listing of steps) and by setting aside time to contemplate their progress and efforts. These reflections can serve as conversation starters—conversations about what they are doing and planning to do and what they did and have learned. I suggest moving toward conversations and notes rather than checklists, rubrics, and more formal evaluations, which seem to distance the student from what she/he is doing, has done, or might do.

These principles stem from a concern that new assessment efforts need to be principled and thoughtful rather than faddish. They reflect a need for a major paradigm shift as regards how we assess, why we assess, and the ways these assessments are manifest in the classroom. Some ramifications include a new type of professionalism on the part of teachers, a shift in the relationship between testing and teaching and between teacher, students, and parents. In general, these principles call for a willingness to recognize complexity and diversity and an approach to assessment that begins from inside rather than outside the classroom. Are we succeeding in terms of shifting such values? Currently, there are several efforts occurring that are simultaneously studying and supporting such shifts (see, for instance, Tierney, Clark, Fenner, Wiser, Herter, & Simpson, in press). I am optimistic enough to think we have the makings of a movement that is beginning to establish its own identity—one that is aligned with contemporary views of learning, and more consistent with pluralistic and constructivist ethics (see especially Moss, 1996).

These principles stem from a concern that new assessment efforts need to be principled and thoughtful rather than faddish.

The principles

Principle 1: Assessments should emerge from the classroom rather than be imposed upon it. Classrooms are places where wonderful ideas are encountered every day; where children engage with one another in a myriad of social interactions; where learning can occur as the culmination of a unit of work, in

conjunction with an experiment, or as students work with others or watch others work. Learnings may be fleeting, emerging, reinforced, and challenged. Oftentimes teachers expect certain learnings; at other times, teachers are surprised at what is learned.

The learnings that occur in classrooms are difficult to predict. Children are different not only in their interests and backgrounds, but also in terms of their literacies. While most teachers may begin the year with a sense of what they want to cover, generally they do not consider their plans to be cast in stone. Indeed, they are quick to adjust to their assessment of their students' needs and even to discard and begin afresh. They are more apt to begin with a menu of possibilities and an open-ended agenda, which allows for learning that is opportunistic and individualized.

With the movement to more child-centered approaches, teaching and learning have become less prescriptive and predetermined and have given way to notions of emergent literacy and negotiated curriculums. Most teachers espouse following the lead of the child. Unfortunately, testing practices tend to abide by a different orientation. Many forms of traditional tests do not measure what is valued and what is occurring in classrooms. Changes in testing have not kept pace with shifts in our understanding of learning and literacy development. Moreover, they often perpetuate an approach to assessment that is from the outside in rather than from the inside out. Indeed, I often argue that one of the reasons for emergent assessment is to ensure that assessment practices keep up with teaching and learning rather than stagnate them by perpetuating the status quo or outdated views of literacy learning.

Compare, if you will, these two scenarios:

Students in one classroom are engaged in a wide array of reading and writing experiences, projects, book talks, conferences, and workshops. In conjunction with these activities the students keep journals in which they discuss their reflections, including their goals and self-assessment of their achievements. In addition, each student maintains a log of his or her reading and writing activities, as well as a folder that contains almost everything. Portfolios, in turn, are used to keep track of the key aspects of their work over time. During teacher conferences with the students, the teacher encourages the students to note what they have achieved and want to pursue further. The teacher keeps her/his own informal notes on what is occurring—focusing on a menu of different aspects drawn from a menu of possibilities that the teacher and some colleagues developed. The menu supports but does not constrain the notes that the teacher keeps on the students. As part of the process, these notes are shared with the students, who are encouraged to add their own comments to them. At parent-teacher conferences and student-led parent conferences both the teacher and the student refer back to these notes, portfolios, etc. to remind themselves of and share what has occurred.

The students in another classroom are engaged in a wide array of activities but are not encouraged to monitor themselves. Periodically the teacher distributes a checklist to each student with a preset listing of skills that the child has to check. Likewise the teacher may interrupt the flow of activities and check the students in terms of these preset listing of skills. The skills on the list bear some relationship to some things that are done, but there are a host of things that are not included and some other things that are included that do not seem to apply. The listing of skills was not developed by the teacher nor is it open ended. Instead, the list was developed by a curriculum committee for the district. In some ways the list reflects a philosophy and approach that do not match the current situation. Nonetheless, the teacher is expected to keep the checklist and file it. After the checklist is completed and filed it is not reexamined or revised.

The first example is representative of an inside-out approach—that is, what is assessed and the manner in which the assessment of various learnings is carried out and originates from within the classroom. An inside-out approach does not involve overly rigid a priori determinations of what should be looked for nor does it restrict the types of learning to be examined. In addition, assessment is negotiated among the parties that are involved.

Our second example may give the illusion of being inside out, but it actually perpetuates the outside-in approach. In this classroom the teacher uses informal assessment procedures, but they do not fit with or emerge from the classroom, and there is no negotiation between

teacher and student. While the second type of classroom may represent an improvement over classrooms that depend upon standardized assessments and periodic checks, it has some major shortcomings in terms of what is being done and how these things are negotiated. Such a classroom does not invest in or trust the professionalism and problem-solving abilities of teachers, as well as the need for student involvement.

Principle 2: Effective testing requires teacher professionalism with teachers as learners. Many of the assessment practices in schools (especially standardized tests) have a dysfunctional relationship with teachers and learners. Whereas in most relationships you expect a give and take, actual testing practices in schools seem more estranged than reciprocal, more detached than intimate. This should come as no surprise for oftentimes testing personnel have separated themselves and their instruments from teachers and students. Testing divisions in school districts generally have detached themselves from teachers and students or have forced teachers and students to work on their terms. In some districts, the testing division may use tenets tied to notions of objectivity and reliability to leverage control of what is tested as well as how, when, and why testing occurs.

If teachers become involved in making assessment decisions, the complexity of dealing with individual differences and differences across classes and schools is apt to surface. It may become problematic to assume that different students can be assessed with the same test, that comparisons across students are straightforward, or that students' performance and progress can be adequately represented with scores derived by periodical administrations of tests.

Quite often teachers will make reference to the tests that they are required to use, principals will allude to the district and state policy, and the district and state lay the responsibility on the public. Some systems seem to be either resistant to change or entrenched in their commitments.

But, teachers relinquishing control of assessment leads to a loss of self-determinacy and professionalism, which is problematic for a number of reasons. It seems to accept and reinforce the view that teachers cannot be trusted. It removes responsibility for instructional decisions from the hands of those who need to be making them. As a result, it decreases the likelihood that assessment will be aligned with teaching and learning and increases the separation between how learning is occurring in classrooms and how it is tested and reported. It depersonalizes the experience and serves as an excuse for relinquishing responsibility. Essentially, the external control of testing and standardization of testing procedures tend to perpetuate teacher and student disenfranchisement.

Teachers are in a better position to know and learn about an individual's development than outsiders.

Teachers are in a better position to know and learn about an individual's development than outsiders. They are with the student over time across a variety of learning situations. As a result they become aware of the subtle changes and nuances of learning within and across individuals. They are sensitive to student engagement, student interests, student personalities, and the idiosyncrasies of students across learning activities. They are less likely to overstate or ascribe too much significance to results on a single test that may have an alienating impact upon a student. They are in a better position to track and assess learning in the context of teaching and child watching, and therefore to help students assess themselves. Effective teachers are effective learners themselves; they are members of a community of learners in a classroom.

So how might assessment be changed? Teachers, in partnership with their students, need to devise their own classroom assessment systems. These systems should have goals for assessment tied to teaching and learning. These goals should be tied to the types of learning and experiences deemed desirable and, therefore, should be established by those most directly invested in the student's education—the teachers and the students them-

selves. These standards/features should be open ended and flexible enough to adjust to the nuances of classroom life. Tied to these goals might be an array of assessment activities from formalized procedures to very informal, from student self-assessment activities to teacher observations to periodical assessments via portfolios or other ways of checking progress.

Teachers and students need to be willing to change and recognize that there exists no quick fix or prepackaged way to do assessment. Indeed, prepackaged assessments are apt to be the antithesis of what should be developed. Unfortunately, teachers, students, and caregivers may have been enculturated to view assessment as predetermined rather than emergent and as having a look and feel quite different from more direct and classroom-derived assessments.

More direct forms of assessment might involve ongoing monitoring of students by sampling reading and writing behaviors, maintaining portfolios and journals, holding periodic conferences, and keeping anecdotal records. Several teachers and state efforts suggest that the community will support, if not embrace, such changes. We have numerous affidavits from teachers to that effect, which are corroborated by published reports of others such as Shepard and Bliem (1995), who found community support for performance assessments or more direct methods of assessment over traditional assessments was forthcoming and considerable when caregivers were presented with examples of the options.

Principle 3: Assessment practices should be client centered and reciprocal. The notion that assessment should empower students and caregivers suggests an approach consistent with a more client-centered approach to learning. A client-centered approach to assessment is not novel. In areas such as psychotherapy and medicine, client-centered orientations are more the rule than the exception. In a court of law the judicial process hinges upon the notion of advocacy for a client. In attempts at being client centered, teachers are apt to consider what students take away from tests or teacher-student conferences. A shift to client-centered approaches addresses how assessment practices are helping students assess themselves—i.e., the extent to which students might know how they can check their own progress. Indeed, the development of assessment practices with such provisions may have far-reaching consequences. It suggests that we should shift the whole orientation of assessment from developing better methods of assessing students toward better methods of helping students assess themselves.

So how might client-centered assessment look? It would look like child-centered learning. Teachers would strive to help students assess themselves. Their orientation would shift from subjecting students to assessment practices to respecting students for their self-assessment initiatives. This entails a shift from something you *do to* students to something you *do with* them or help them *do for themselves*—a form of leading from behind.

A number of classrooms have in place the beginnings of student self-assessment vehicles via the use of journals, logs, and portfolios. But this is just a beginning; self-assessment should extend to every aspect of the classroom, from helping students formulate their own learning goals, to helping students make decisions on what they can handle and need, to having them collaborate in the development of report cards and parent-teacher conferences. Too, the involvement of students in their own assessment helps with the management of such activities. This might entail having students set their own goals at the beginning of a unit (not unlike what is proposed with K-W-L); hold conferences with teachers, parents, or peers as they progress or wrestle with issues; look at their efforts and study their progress; and set future goals at the end of a unit in conjunction with parent conferences, or as alternatives to report cards.

There are numerous ways to start these conversations. I ground my conversations about assessment for and with students in the actual portfolio without the intrusion of a grade or score. Scores and grades only give the illusion of accuracy and authority; conversations connected to portfolios or other forms of more direct assessment unmask the bases for decision making and spur the conversation toward a consideration of the evidence, an appreciation of assumptions and the negotiations of goals. "Let's look," "I can show you," "It's like this," "I see what you mean," and "Do you think" displace more general and removed

conversations, which tend to be categorical rather than contributory.

Various forms of self-analysis can complement portfolios and be wonderful springboards for such conversations. For example, sometimes I will have students represent their progress and goals with bar graphs or other visual representations (e.g., Venn diagrams, landscapes) in a fashion akin to "then," "now," and "future" and use these graphs as conversation starters. In turn, the visuals serve as the basis for having students delve into their portfolios and examine evidence about what they have achieved and what they might focus upon or set their sights on.

Principle 4: Assessment should be done judiciously, with teachers as advocates for students and ensuring their due process. A useful metaphor, if not rule, for rethinking assessment can be derived from aligning assessment with judicial processes. In a court of law, an individual on trial is given an advocate who presents evidence, including testimony, to present a case on behalf of the client. The client and the lawyer work in tandem. The trial is judged upon whether or not the client was given a just hearing and whether or nor her or his representation was adequate. The client has the right to see the evidence presented for and against her or him, the right to reports developed, the right to present his or her own evidence and arguments, and the right to appeal. Also, in the event the client is not satisfied with his or her representation, the client has the right to request someone else to support his or her making a case or, if concerned about procedure, to request a retrial.

Now consider how students are put on trial in our school systems. They may or may not have an advocate, they may or may not be given adequate representation, and the evidence that is presented may or may not best represent their cases. They may not see the reports that are developed. Indirect indicators such as standardized tests, of questionable (if not circumstantial) quality, serve as the basis for decisions that restrict opportunities. In a host of ways assessment activities appear less judicious than they should be. Indeed, students are rarely given the right to appeal or to provide their own evidence—it is as if the students' right to due process is violated.

An examination of the law governing public schools raises some interesting concerns regarding schooling. Over the last 30 years, some key U.S. Supreme Court decisions have been offered that should direct our thinking. In Tinker v. Des Moines Independent School District, 393 U.S. 503 (1969), a case involving freedom of speech, the Court established some key principles undergirding students' rights. The Court wrote: "In our system, state operated schools may not be enclaves for totalitarianism....Students in schools, as well as out of school are possessed of fundamental rights which the State must respect." This position was reaffirmed in the case of Goss v. Lopez, 419 U.S. 565 (1975). As Justice White stated, "young people do not 'shed their constitutional rights' at the schoolhouse door"—the right to due process is of particular importance when the impact of an event "may interfere with later opportunities."

I would hope that legislators pursue practices that place students' rights at a premium rather than displace such a goal with practices that serve first to protect themselves against legal challenges. At a minimum, I would hope that any assessments afford students better due process, including the right of disclosure and presentation of evidence on behalf of the student, as well as the right to appeal the use of indirect or circumstantial evidence. Moreover, I would hope my appeal for judicious assessment shifts the pursuit of such to being both a goal and a right.

Unfortunately, some U.S. state legislators may be more intent on protecting themselves against possible litigation than ensuring that students' rights have been fully supported. For example, they might consider that the spirit of due process has been satisfied when students have been given advance notice of tests and what these tests will entail—that is, in lieu of opportunities to appeal or students providing their own "alternative" evidence of progress or proficiencies. Also, an insipid development occurs when teaching to the test is used to maximize the legal defensibility of tests. In particular, states will often try to finesse the possibility of legal challenges of test bias by ensuring that students have had the opportunity to learn the content covered on tests. To avoid litigation and appear to address local needs, they will establish programs to prepare students for the tests and therefore

"make" their tests unbiased by definition. The attitude of most institutions and states is to emphasize legal defensibility ahead of protection for and advocacy on behalf of students.

Principle 5: Assessment extends beyond improving our tests to the purposes of assessment and how results from assessment are used, reported, contextualized, and perceived. Any consideration of assessment needs to be broadly defined to encompass an exploration of the relationship between assessment and teaching, as well as facets such as report cards, parent-teacher-student conferences, and the student's ongoing record. These facets should not be viewed as exempt from scrutiny in terms of the principles described herein. They should be subjected to the same guidelines.

Rather than keep the parent or caregiver at arm's length in the negotiations over reform, we need to embrace the concerns that parents have and the contributions that they can make.

Just as the goals for developing better classroom-based assessment procedures are tied to the principles discussed herein, so report cards, records, and other elements must be examined in terms of whether they adequately serve the ends for which they are intended. Take, if you will, report cards. Do report cards serve the needs of the student, teacher, and parent? Do they represent a vehicle for ongoing communication and goal setting? Are they done judiciously? If not, how might the method of reporting be changed to afford such possibilities? Or, take, if you will, the student's records. For what purposes are the records used? Are the records adequate for these purposes?

Changes in assessment should be viewed systemically. When teachers contemplate a shift in classroom assessment, it is rarely a matter of simply making selected adjustments or additions. What a teacher does with one facet should and will affect another. For example, a teacher who incorporates a portfolio approach is likely to become dissatisfied with traditional forms of reporting progress. The solution is not to shy away from such changes, but to realize that they will need to occur and, if they do not, to realize that the failure to make such changes may undermine the changes already made. Teachers start to feel as if their new assessment initiatives are being compromised. Students may begin to sense mixed messages if teachers advocate student decision-making and then reassert their singular authority via the determination of a grade without any student input or negotiation. That is, teachers move in and out of assessment practices tied to very different underlying principles. I feel as if the worth of assessment efforts such as portfolios may be diminished if the portfolios are graded or graded inappropriately either without any student input or without consideration for diversity and richness—especially, what the portfolio might mean to the student. We need to keep an eye on achieving students' engagement in their own learning as we negotiate future goals and possibilities against the type of judgments that are made and reported by whom and how.

We should not underestimate the importance of parent or caregiver involvement in such efforts. Rather than keep the parent or caregiver at arm's length in the negotiations over reform, we need to embrace the concerns that parents have and the contributions that they can make. In those situations where teachers pursue alternatives to report cards, parent contributions may be crucial. Parents need to be informed of the goals and engaged in contributing to the efforts. Because not all parents might see the advantages, they may need choices. And, there are ways to avoid holding all parents hostage to what one parent or a small number express as concerns. For example, in pursuit of student-led conferences as an alternative to report cards, Steve Bober (1995) presented parents in Massachusetts with a description of two alternatives and offered them a choice—student-led conferences or more traditional report cards. Parents choosing student-led conferences were also expected to write letters to their children after each conference. Apart from the distinctiveness of the

practice, what is notable is how Bober engaged parents as informed partners in the practice.

Principle 6: Diversity should be embraced, not slighted. Oftentimes those assessing students want to remove any cultural biases rather than recognize diversity and support individual empowerment. They often pursue culture-free items and analysis procedures as a way of neatening and comparing. In pursuit of straightforward comparisons they assume that to be fair more items are needed, and therefore, the use of authentic assessment procedures will create problems, especially since the "time-consuming nature of the problems limits the number" (Linn, Baker, & Dunbar, 1991, p. 18). In addition, they seem to support as a given the use of the same analysis systems for the responses of all students. They expect a respondent to interpret a task in a certain way and respond in a set manner and may not tolerate variation in response, even if such variation might be justified. Whereas they might allude to the context-specific nature of any assessment, they tend to retreat from considering individuals on their own merits or in their own ways.

The term *culture-free tests* seems an oxymoron. I suspect that it is well nigh impossible, and certainly questionable, to extract cultural influences from any test or measure of someone's literacy. Literacy, your own and my own, is inextricably connected to cultural background and life experiences. Culture-free assessments afford, at best, a partial and perhaps distorted understanding of the student. In other words, assessments that do not build upon the nature and nuances of each individual's experiences should be viewed as limited and perhaps flawed. Just as teachers attempt to engage students by building from their background of experiences, so assessment should pursue a goal of culture sensitivity. Classroom teaching does not occur by ignoring or removing diversities. Nor should such a view of assessment be dismissed because of its ideological or sociopolitical considerations: Recognition or validation of one's own experience would seem a basic human right.

We need to aspire to culturally based assessment practices. In some ways I see this pursuit consistent with John Ogbu's (1988, 1991) notions about beginning to meet the needs of African American students—namely, an approach to educational reform that has a cultural ecological orientation. I envision cultural ecological assessments that build upon, recognize, and value rather than displace what students have experienced in their worlds.

For a number of years literacy educators have been willing to sidestep complex issues of culturally sensitive assessments by appealing to the need to make straightforward comparisons. For years standardized test developers and the National Assessment of Educational Progress have retreated from dealing with issues of nonuniformity and diversity as they have pursued the development of scales for straightforward comparisons across individuals. In conjunction with doing this, they have often revised their assessment instruments to ensure that results fit their models or views of literacy. For example, they are apt to exclude items on topics tied to specific cultural interests and to remove items that show an advantage for one group over another. Even recent attempts espousing guidelines for new approaches to performance assessment (e.g., Linn et al., 1991) or exploring bias in testing minorities (Haney, 1993) may have fallen prey to the same view of the world.

Principle 7: Assessment procedures may need to be nonstandardized to be fair to the individual. As teachers try to avail students of every opportunity within their control, they are constantly making adjustments as they "read" the students—their dispositions, verbal abilities, familiarities, needs, and so on. We look for ways to maximize the learning for different students, and we know that different students may need different amounts of encouragement and very different kinds of support. If we standardized our teaching, we know what would apt to be the end result—some students with wonderful potential would reveal only certain sides of themselves and might not achieve their potential or even reveal who they are and what they might contribute and learn.

Allowing for individual or even group differences creates havoc with the desire to standardize assessment. Standardization approaches each individual and group in the same way—that is, students perform the same tasks at the same time, and then their responses are assessed using the same criteria. But if different students' learning repertoires are different and different students enlist different strategies and

have different values, etc., and different approaches to testing, then what may be standard for one student may be unique for another.

Studies across cultures, across classrooms, and within classrooms suggest that different students respond in different ways to different forms of assessment depending upon their histories—cultural, classroom, or personal. As my previous principle suggested, how students respond should be looked at as different across situations and against a "comparative canvas, one that takes into account the nature of the community that students inhabit, both the community of the classroom and the community of society with all of its past and present conditions and hopes for the future" (Purves, 1982, p. 345). Green and Dixon (1994) have emphasized that students construct "situated repertoires associated with particular models for being a student...not generic ones" (p. 237). We have ample demonstrations as to how the responsiveness of various groups and individuals in testing situations depends on their view of the social dynamics of the situation (Basso, 1970; Crumpler, 1996; Ogbu, 1988; Philips, 1983).

Indeed, there is always a tension between a need for uniformity across individuals and groups and the use of procedures that are sensitive to the different literacy developments of students, as well as the students' own predispositions to respond differently to different people in different ways at different times. On numerous occasions my assessment of some students has been revised as a result of pursuing more than one mode of response, as well as establishing different kinds of partnership with them or watching them interact over time in different situations with different individuals or groups. In turn, what may serve as a vehicle for uncovering the literacies of one student may not be a satisfactory method for uncovering those of another student or those of the same student at another time. Teachers need to be willing to use different means with different students whether they are assessing or teaching.

The decision-making process may also be complicated by certain of our own predilections. In conjunction with my work on portfolios, I am always surprised at the analyses that learners have done of their progress and the types of goals that they choose to pursue. They ascribe to elements in their portfolios significance that I may have overlooked or not have been able to see. And, their decisions to proceed are often at variance with what I would have suggested.

Principle 8: Simple-minded summaries, scores, and comparisons should be displaced with approaches that acknowledge the complex and idiosyncratic nature of literacy development. Straightforward comparisons across individuals are usually arbitrary, biased, and narrow. Assuming an approach to assessment with a new openness to complexity, respect for diversity, and interest in acquiring a rich picture of each student, then how might decisions be made about students? Those decisions that require reflection upon the individual's progress and prospects will likely be bountiful. Teachers who pursue an open-ended and diverse view of students will find little difficulty negotiating new areas of pursuit with and for individual students. Decisions that demand comparisons from one individual to the next will be problematic, but these difficulties are not insurmountable. They require a willingness to deal with uncertainties, to entertain possibilities, and to negotiate decisions, including the possibility that there will be lack of agreement. The problems with comparisons are confounded when people assume that straightforward continuums or single scores can adequately describe students.

Comparisons based upon scores are so problematic for a host of reasons: (a) Each student's development is unique; (b) the literacies of one student will be different from another, and even the same literacies will involve differing arrays of facets; and (c) some of these facets will be unique to a certain situation. Literacy development is sufficiently different from one student to the next that the types of comparisons that might be made are quite complex and multifaceted. The term *literacy abilities* rather than *literacy ability* seems in order. If you were trying to portray the character of these developments, you might find yourself gravitating to describing individuals on their own terms. Unfortunately, the terms of comparison in place with standardized tests and NAEP assessments and implicit in many of the attempts to score portfolios and other classroom-based data are often insensitive to such complexity. Looking at different individuals in terms of a single score masks variability and in-

dividuality. Again, test makers err on the side of a level of simplification not unlike a massive "conspiracy of convenience" (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987, p. 180).

The drive for uniformity is quite pervasive. Our assessment and instructional programs oftentimes include long lists of skills as outcomes to be assessed, taught, and mastered. It is assumed that skills are neatly packaged and discrete and that each makes a uniform contribution to literacy development. It is assumed that students acquire these skills to mastery and that their ability to use them is uniform across literacy situations. In authentic reading and writing situations within which genuine purposes are being pursued, this is unlikely. Across literacy situations certain attributes may be more likely to be enlisted than others, and they are apt to be enlisted as clusters rather than one by one or discretely.

Too often literacy educators have ignored the complexities of the issues and have fallen back on convenience rather than exploring possibilities. Take, if you will, the attempts to wed some of the data emerging from performance assessment (e.g., portfolios) with rubrics. The data generated from a portfolio might involve a rich array of samples or observations of the students' work across situations and time. These samples are apt to represent the students' pursuit of different goals, utilizing different resources, including content, under varying conditions. In some ways student classroom samples may vary as much as the works of art from an artist's portfolio. Each sample may represent very different achievements and processes. When you hold them, examine them, and discuss their significance you are in touch with the actual artifact and not some distant derivative.

It is at this point, some would argue, that we can use a rubric to affix a score or scores or a sum total score to the student's work. But we need to examine a question that is the reverse of what is often asked. Instead of asking how we rate the portfolio, we should be asking whether the rubric measures up to the portfolio or to the assessment of complex performance. Moreover, in classrooms do we need a measure that is a distant derivative when we have the primary sources—the actual samples—to examine and reexamine using an array of lenses or perspectives? Whereas I argue for the context-specific nature of any assessment, advocates of rubrics seem to want to dismiss idiosyncrasies and variation—that is, they would retreat from being willing to consider individuals on their own merits or in their own ways. Unless rubrics are used to prompt a consideration of possible ways to analyze work or as conversation starters in conjunction with revisiting the students' work samples, I see few advantages to their use in classrooms.

Sometimes assessment of reading and writing becomes more far-fetched by adding together a set of subscores. A key assumption often undergirding the use of such scores—especially the suggestion that they can be added and used as the basis of comparative decision making—is that the full and detailed portrait of an individual's literacies has been afforded. Unfortunately, these dimensions are not exhaustive, these determinations of degree are not accurate, and they should not be added. To be able to do so, we would have to do the following:

1. include all of the attributes or be assured that the partial listing that was developed is representative;
2. determine how these attributes are configured across situations;
3. assume that ample evidence will be provided for assessing these attributes;
4. develop scales for assessing attributes; and
5. generate an algorithm that works across individuals by which we might combine the elements and their dimensions.

I would posit that we do not have such samples, sampling procedures, ways of procuring evidence, adequate scales, or algorithm. And it is problematic to assume that an algorithm that simply represents sums would ever be adequate. The complexity of literacy is such that we cannot assume a basis for generating or combining scores.

Literacy assessments cannot and should not be so rigid. Perhaps there are some benchmarks that are appropriate across all students. Perhaps there are benchmarks appropriate to some readers and not others. But such benchmarks are likely to represent a partial view of any student's literacies. The use of scores and continua as ways of affording simplification and comparability has a tendency to camouflage the subjectivity of assessment and give

test developers the allusion of objectivity. The use of scores and continua is not more objective; it is arbitrary. Guba and Lincoln (1989) have suggested the shift toward accepting the inevitability of relativism and the complexities across different settings may require the ongoing, ecumenical, and recursive pursuit of shared possibilities rather than a single set of absolute truths.

Principle 9: Some things that can be assessed reliably across raters are not worth assessing; some things that are worth assessing may be difficult to assess reliably except by the same rater. Oftentimes, test makers and researchers will perseverate on whether or not they can consistently measure certain abilities. They tout reliability as the major criteria for whether or not a test is valid. The end result is that some things that are worth measuring are discarded and some things that are not worth measuring or valuing achieve an elevated level of importance. Typically, complex and individualistic learning tends to be shortchanged whereas the currency of learnings that are easier to define may be inflated. For example, in writing assessment, constructs such as style or voice may be shortchanged, while spelling and punctuation may be inflated. In reading, constructs such as self-questioning, engagement, and interpretation may be shortchanged, while speed, factual recall, and vocabulary may be elevated.

Unfortunately, reliability is translated to mean that two different scorers or raters will be able to assess the same thing in the same way. Unless a high degree of agreement across raters is achieved, test makers will deem a measure unreliable and therefore question its worth. In so doing, they may be making the mistake of assuming that reliability equates to agreement when verifiability may be a better approach.

We should be willing to accept differences of opinion in terms of how certain abilities are rated or discerned. Some abilities and strategies are difficult to pin down in terms of clear operational definitions. Different raters or even the same raters at different times are apt to develop different constructions of the same phenomena. Sometimes these shifts arise as a result of the different predispositions of the raters. Sometimes they arise as different facets of the phenomena are taken into account either by different raters or the same rater. Sometimes they arise as a result of differences in how students enlisted certain abilities. Such differences should not be viewed as surprising, for they coincide with two key tenets of most current views of learning: the notion of an ongoing constructive nature of knowing; and the situation-specific nature of learning. Differences are apt to exist across and within an individual's literacies (e.g., reading a newspaper for purposes of locating an advertisement versus reading a romance novel for pleasure) and from one individual to the next. In other words, some features may or may not apply to some students' literacy, and some facets may apply uniquely to individuals.

One should not be seduced into thinking that variables that are easy to define should be looked at to the exclusion of those that are difficult to assess. It may be foolish to exclude some facets because they are difficult to assess or because they look different either across students or situations or by the raters. Likewise, one should not be seduced into thinking that every reading and writing act is the same and involves the same variables. If the only literacy facets scored are those common across students and those that can be scored with high reliability across different students' responses, then certain facets will be given more weight than they deserve, and some important facets may be excluded.

Principle 10: Assessment should be more developmental and sustained than piecemeal and shortsighted. To assess how well a student is doing, our vision or vistas need to change. If assessment goals are tied to development, then we need to look at patterns and long-term goals. What we see or look for in a single selection or case may not be helpful in looking for patterns across cases, selections, or circumstances. For example, as a reader or writer reads and writes a single selection, we might look for engagement and active involvement. Across situations we might want to consider the extent to which the interest and engagement are maintained across a range of material for different purposes. We also might be interested in the extent to which the student has developed a value for reading and writing that is reflected in how he or she uses reading and writing inside and outside the class. This may be apparent in her or his self-selection of

Short- and long-term contrasts in assessment

Short-term/single instance	Long-term/multiple situations
Affect	
Engaged	Value
Active	Self-seeking ongoing
Thoughtful	Habit
Strategies	
Planning	Flexible, reflective, coordinated, selective, customized
Fixing up and troubleshooting	
Making connections	
Looking back, forward, and beyond	
Collaborating	Community building
Outcomes	
Main idea	Overall understandings, intertextual connections
Details	Projects
Conclusions	Applications
Implications	Range of problems and activities
	Overall understandings and themes
Self-assessment	
Self-monitoring	Self-scrutiny, goal setting, self-determinations
Online problem solving	Overall goals, progress, patterns

books or self-initiated writing to serve different purposes.

Within areas such as the students' abilities to read with understanding, our goal for a single selection might be the extent to which a reader understands the main idea or theme or can draw conclusions using selected details, etc. Across selections or in the long term, we might be interested in how the students use different books to contribute overall understandings tied to units or projects or their own developing understandings of the world. Or, we might be interested in self-assessment. With a single selection we could focus on the reader's or writer's ability to monitor reading and writing, to set goals for a specific selection, and to problem-solve and wrestle with meaning-making. Across selections we might be interested in the reader's or writer's ability to set goals and assess progress across several selections. In looking across selections, you should not expect that students will always appear to reveal the same level of sophistication with skills and strategies or necessarily use the same skills and strategies. See the Table for other short-term, long-term contrasts.

A shift toward assessment that examines students over time aligns assessment with classroom practices that pursue sustained engagement and aim to help students derive an understanding of patterns. It shifts our teaching and learning to long-term possibilities rather than the specific and short-term objectives of a lesson.

Principle 11: Most interpretations of results are not straightforward. Assessment should be viewed as ongoing and suggestive, rather than fixed or definitive. In many ways teaching involves constant redevelopment or continuous experimentation and adjustments to plans, directions, and future goals. To appreciate the complexities and sophistication of teaching, consider the image one conjures up for a sportsperson. In certain sports (e.g., baseball, tennis) involving eye-hand coordination with racquets or bats, players will begin their swing and constantly be making subtle adjustments as balls with different velocities, rotations, and angles are thrown at them. But sporting events pale in comparison with the dynamics of teacher-student interactions—the adjustments, just in time decision making, and ebb and flow of activities that occur. Teachers deal with students whom they may be trying to respond to, motivate, mobilize, develop, and coach while understanding

their needs, beliefs, strategies, and possible ways of responding as they are interacting with one another and dealing with the rest of their lives. Not surprising, teachers have to be a mix of ecologist, developer, advocate, coach, player, actor-director, stage manager, mayor, and sometimes counselor. Teachers are always planning and recognizing the need to make constant adjustments to what they are doing and what they might do next.

For these purposes, the typical assessment data (e.g., scouting reports of students provided by school records, premeasures of abilities, standardized or even informal assessments) may provide limited guidance to teachers in terms of the moment-by-moment decision making and even planning for the next day or week or even month. Too often typical student records seem as limited as a mug shot taken of the learner; you may be able to identify the learner (depending upon your ability to see likenesses) but may not. Certainly, the mug shot will not afford you an appreciation of the character of the student, nor will it help you understand the range of things that the student can do, nor will it support your ability to negotiate either long-term or short-term learning goals.

Most classroom-based assessments offer more promise but are still limited. Classroom-based assessment procedures may give teachers a better sense of how students will proceed in like circumstances and may also afford a fuller picture of the student across time. Portfolios, for example, are equivalent to scrapbooks involving multiple snapshots of the learner in a variety of contexts. Such assessments might afford a fuller and richer depiction of the learner and his or her pattern of development, but judgments—especially prescriptions—are never as straightforward as they might appear. The possibility of obtaining a complete vision of a learner is complicated by our inability to constantly monitor a learner, delve into and interpret his or her innermost thoughts, and achieve more than one perspective on the learner. It is also tied to the ever-changing nature of learning. Apart from the fact that our snapshots of classroom learning tend to be still shots of the learner, these images are tied to a place and time that has become more historical than current. Such limitations might be viewed as a problem if we were to perseverate on wanting to pin down what to do next with a student and be sure to stick to a set course. Instead, they should be anticipated and viewed as tentative bases for where and when one might begin. While we can develop short- and long-term goals and plans, we should not approach our teaching as if our prescriptions should not be altered, assessment fixed, nor directions more than suggestive.

Likewise, we should not approach assessment as if our results need be final or base our subsequent actions as if we have derived a decision that is any better than a hunch. We should avoid assuming that our assessments do anything more than afford us information that we might consider. No assessment should be used as restrictively or rigidly as decisions made in courts of law, yet I fear that many are. Instead we should reinforce what needs to occur in classrooms—constant adjustments, shifts, and ongoing decision making by teachers who are constantly watching, learning, coaching, and responding to students, peers, and others.

Principle 12: Learning possibilities should be negotiated with the students and stakeholders rather than imposed via standards and assessment that are preset, prescribed, or mandated. The state within which I reside (along with many other states) has been seduced into thinking that standard setting may be the answer to improving education by ensuring that teachers teach and students learn certain basic skills. I find myself quite discouraged that our professional associations have aligned with similar efforts. Historically, standard setting (and the proficiency testing that it spurs) has tended to restrict access and experimentation at the same time as it has tended to support agendas tied to gatekeeping and exclusion.

The standard-setting enterprise and the proficiency-testing industry have the potential to perpetuate the view that we can set targets that we can easily reach. Unfortunately, it is problematic to assume that development is simply setting a course for the student from A to B—especially when A is not taken into account and B is tied to views of outcomes looking for expertise rather than individual assessment of development. Without ample consideration being given for where students are and how and why they develop and their aspirations, we are apt to

have our targets misplaced and our learning routes poorly aligned. I was in attendance at one of the many sessions on standards sponsored by the International Reading Association and the National Council for Teachers of English, when a speaker talked about standards using the analogy of a basketball player of the caliber of Michael Jordan as the "standard." As the speaker discussed the worth of setting standards based upon what we view as aspirations, I mulled over my height and my skill and what I might do to improve. Then I reminded myself of my reasons for playing basketball and where I am insofar as my background in basketball. I play basketball for fun, to be with my sons, and for exercise. We need to realize that we should be asking who is deciding? Whose standards are being represented? In some ways the quest for educational improvement via standards and in turn proficiency testing places a premium on uniformity rather than diversity and favors prepackaged learning over emerging possibilities.

In a similar vein, advocates of standards emphasize the importance of the role of making judgments by comparisons to Olympic skating and other activities where success is measured by the trophies one achieves or the graded measures that are applied. I think we need to challenge this metaphor and question the emphasis on judgment rather than support. I prefer to think of a teacher as a coach rather than a judge—a supporter and counselor versus a judge and award- or grade-giver. I would like to see teachers view their role as providing guidance, handholding, and comments rather than As, Bs, and Cs or some score. In my view of a more ideal world, I see teachers, students, and caregivers operating in a kind of public sphere where they are part of the team negotiating for a better self. In this regard, I find myself fascinated with several classroom projects: with the kind of self-reflection and analysis occurring amidst the community-based preschool efforts of Reggio Emilia (Forman, 1993, 1994) where teachers, students, and community work together developing and implementing curriculum plans, ponder the right questions to ask to spur students' reflections, develop insights, and learn; with the work of Short, Harste, and Burke (1996) on developing inquiry in Indianapolis schools (as they engage students and teachers in considering the anomalies, patterns, and ways of looking at themselves); with the work of the Santa Barbara Classroom Discourse Group (1992a, 1992b), a community of teachers, researchers, and students interested in understanding how life in classrooms is constructed and how expectations and practices influence opportunities to access, accomplish, and learn in school; and with the work of Fenner (1995) who uses a general form of Toulmin's (1958) analysis of argumentation to examine classroom conversations and student self-assessments with portfolios and looks for ways to help students look at themselves in terms of evidence, assumptions, claims, and goals. Fenner's approach to self-assessment moves us away from the typical checklist that asks students to detail in rather vague and unsubstantiated fashion their strengths and goals in a kind of "hit and miss" fashion.

Unfortunately, rather than language that suggests a view of classrooms as developmental and nurturing, oftentimes the metaphors adopted by those involved in the testing, proficiency, and standards enterprises seem more appropriate to developing consumer products connected to prescribed guidelines and uniform inspection procedures. That is, they seem to fit with our views of industry rather than nurturing human potential (Wile & Tierney, 1996). With this in mind, I would suggest that we should assess assessment based on whether it is parsimonious with a society's bill of rights and our views of individual rights, opportunities, and freedoms.

I fear that standards will perpetuate the effects uncovered when Ellwein, Glass, and Smith (1988) surveyed the history of the effects of various statewide proficiency testing—gatekeeping and the removal rather than enhancement of opportunities. Indeed, in Ohio and I would suspect other states, Ellwein et al.'s (1988) findings are being replicated. With the introduction of proficiency testing more students are dropping out. Ironically, the tests were intended to improve instruction, but fewer students are taking them, which in turn suggests that more students are passing them. So by keeping these dropouts invisible, advocates of proficiency testing and legislators claim the reform is having positive effects—that is, as more students leave or drop out, abhorring or deterred by the situation, legislators and advocates (in-

cluding the media) erroneously suggest or advertise falsely that more students are passing.

Closing remarks

My principles for assessment emanate from a mix of child-centered views of teaching, developmental views of children, constructivist views of knowing, critical theoretical views of empowerment, and pluralistic views of society. I view them as suggesting directions and guidelines for thinking about the why, how, where and when, who, and what of assessment.

Why?
- To develop culturally sensitive versus culturally free assessments
- To connect assessment to teaching and learning
- To connect assessment to students' ongoing goal setting, decision making, and development
- To become better informed and make better decisions
- To develop assessment that keeps up with teaching and learning

How?
- Collaborative, participatory, client centered
- Coach-like, supportive and ongoing rather than judgmental, hard-nosed, and final
- Supplemental and complementary versus grade-like and summative
- Individually, diversely, not prepackaged
- Judiciously
- Developmentally
- Reasoned

Where and when?
- Amidst students' lives
- Across everyday events and programs
- In and out of school
- Opportunistically, periodically, continuously

Who?
- Students, teachers, and stakeholders

What?
- Ongoing learning: development, resources, and needs
- Complexities
- Individuals and groups
- Evidence of progress and decision making
- Programs, groups, individuals

In describing the essence of my proposition, I would like to return to where I began. I believe an overriding principle, which is perhaps my 13th or more of a penumbra, is *assessment should be assessed in terms of its relationship with teaching and learning, including the opportunities learners are offered and the rights and respect they are accorded.*

Shifts in my own thinking about assessment began occurring when I asked myself this question: If I were to assess assessment, what criteria might I use? My answer to this question was that assessment practices should empower teachers, students, and their caregivers. In other words, assessment practices should enrich teaching and learning. As I explored how tests might be used as tools of empowerment for teachers and learners, I became interested in whether this type of assessment actually helped teachers and students (as well as the student's caregiver, resource teachers, principal, and others) achieve a more expanded view of the student's learning. I also wanted to know whether testing contributed to developing goals and formulating plans of action, which would suggest that assessment practices were empowering. My view of empowerment includes:

> Teachers having a fuller sense (expanded, refined, different) of the students' abilities, needs, and instructional possibilities;
>
> Students having a fuller sense of their own abilities, needs, and instructional possibilities;
>
> Teachers integrating assessment with teaching and learning (this would entail the dynamic/ongoing use of assessment practices, as well as assessment tailored to classroom life); accommodating, adapting, adjusting, customizing—shifting assessment practices to fit with students and their learning and adjusting teaching in accordance with feedback from assessments;
>
> Students engaging in their own self-assessments as they set, pursue, and monitor their own goals for learning in collaboration with others, including peers, teachers, and caregivers.
>
> Communities of teachers, students, and parents forming and supporting one another around this assessment process.

The use of standardized tests, tests accompanying the published reading programs, and even teacher-made tests do not expand teachers' views of their students' learning over time, nor suggest ways the teacher might help them. Nor are such tests integrated into classroom life. They tend to displace teaching and learning activities rather than enhance them.

Likewise, students rarely seem to be engaged in learning how to assess themselves. When my colleagues and I interviewed teachers with whom we began working in assessment 10 years ago, most teachers did not conceptualize the goal of testing to be helping students reflect or obtain feedback on their progress, nor did they envision tests as helping students establish, refine, or achieve learning goals. When we interviewed students, we found that students in these classes tended to have a limited and rather negative view of themselves, and they had set few learning goals. Attempts to examine the impact of more learner-based assessments yielded quite contrasting results. In classrooms in which portfolios were becoming an integral part of classroom life, teachers and students had developed a fuller sense of their own abilities (Carter, 1992; Carter & Tierney, 1988; Fenner, 1995; Stowell & Tierney, 1995; Tierney, Carter, & Desai, 1991).

A study by Shavelson, Baxter, and Pine (1992) provides other confirmation of the worth of aligning assessment to the teaching and learning in classrooms. In their attempts to examine variations in instructional programs, they concluded that direct observations and more emergent procedures captured the shifts in learning while traditional methods (multiple choice, short answer) did not. Such findings should come as no surprise to those of us who have been involved in research on the effects of teaching upon learning; that is, very few literacy researchers would rely upon a standardized test to measure the effectiveness of particular teaching strategies with different students. Instead, we are apt to pursue a range of measures, and some of us would not develop our measures a priori. In fact, several efforts have demonstrated the power of new assessment approaches to evaluate and guide program development and teacher change effectively (see Tierney et al., 1993).

Designing these new assessment approaches has to do with a way of teaching, testing, and knowing that is aligned with a set of values different than what has been and still is espoused by most educational reformers. Unfortunately, the power of some of the psychometricians and their entrenched values related to testing make the emergence of alternative assessment procedures difficult. Indeed, I see the shift as involving a cultural transformation—a shift away from what I view as a somewhat totalitarian practice tied to "old science" and metaphors that equate student learning to quality control.

Mike Rose (1995) suggests in *Possible Lives* that classrooms are created spaces, and the successful ones create spaces where students feel safe and secure; they are the classrooms in which students are willing to stretch, take risks, and pursue their interpretive authority for themselves and with others. In a similar vein, Kris Gutierrez and her colleagues (Gutierrez, Rymes, & Larson, 1995), in discussing teacher-student discourse, assert the need for spaces where students and teachers can connect or transact with each other, rather than pass by one another. The key is finding ways to effect involvement and transaction rather than detachment and monolithic responses.

Assessment must address making futures possible and pursuable rather than impossible or improbable. We must create spaces where students, teachers, and others can achieve futures and spaces wherein the dynamics and practices are such that they challenge but do not undermine the ecology of who students are and might become.

Tierney has done considerable work with teachers and students on assessment issues from portfolios to report cards to studies of teacher and student change. He is currently Professor and Director, School of Teaching and Learning, The Ohio State University, 1945 N. High Street, Columbus, OH 43210, USA.

References

Baker, A., & Luke, A. (1991). *Toward a critical sociology of reading pedagogy*. Philadelphia: John Benjamin's.

Basso, K. (1970). "To give up on words": Silence in Western Apache culture. *Southwest Journal of Anthropology, 26*, 213 – 230.

Bober, S. (1995, July). *Portfolio conferences*. Presentation at Lesley College Literacy Institute, Cambridge, MA.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Carter, M. (1992). *Self-assessment using writing portfolios*. Unpublished doctoral dissertation, The Ohio State University, Columbus.

Carter, M., & Tierney, R.J. (1988, December). *Writing growth: Using portfolios in assessment*. Paper presented at the National Reading Conference, Tucson, AZ.

Ellwein, M.C., Glass, G.V., & Smith, M.L. (1988) Standards of competence: Propositions on the nature of testing reforms. *Educational Researcher*, *17*(8), 4 – 9.

Fenner, L. (1995). *Student portfolios: A view from inside the classroom*. Unpublished doctoral dissertation, The Ohio State University, Columbus.

Forman, G. (1993). Multiple symbolizations in the long jump project. In C. Edward, L. Gandini, & G. Forman (Eds.), *The hundred languages of children* (pp. 171 – 188). Norwood, NJ: Ablex.

Forman, G. (1994). Different media, different languages. In L. Katz & B. Cesarone (Eds.), *Reflections on the Reggio Emilia approach* (pp. 41 – 54). Urbana, IL: ERIC/EECE.

Freire, P., & Macedo, D. (1987). *Literacy: Reading the word and the world*. South Hadley, MA: Bergin & Garvey.

Gee, J. (1990). *Social linguistics and literacies: Ideologies in discourse*. New York: Falmer Press.

Green, J., & Dixon, C. (1994). Talking knowledge into being: Discursive and social practices in classrooms. *Linguistics and Education*, *5*, 231 – 239.

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Gutierrez, K., Rymes, B., & Larson, J. (1995). Script, counterscript, and underlife in the classroom: James Brown versus Brown v. Board of Education. *Harvard Educational Review*, *65*, 445 – 471.

Haney, W. (1993). Testing and minorities. In L. Weis & M. Fine (Eds.), *Beyond silenced voices* (pp. 45 – 74). Albany, NY: State University of New York Press.

hooks, b. (1989). *Talking back*. Boston: South End Press.

hooks, b. (1994). *Teaching to transgress: Education as the practice of freedom*. New York: Routledge.

Lather, P. (1986). Research as praxis. *Harvard Educational Review*, *56*, 257 – 277.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex performance assessment: Expectations and validation criteria. *Educational Researcher*, *20*(8), 15 – 21.

Moss, P. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, *25*(1), 20 – 28.

Ogbu, J. (1988). Literacy and schooling in subordinate cultures: The case of Black Americans. In E. Kintgen, B. Kroll, & M. Rose (Eds.), *Perspectives on literacy* (pp. 227 – 242). Carbondale, IL: Southern Illinois University Press.

Ogbu, J. (1991). Cultural perspective and school experience. In C. Walsh (Ed.), *Literacy as praxis: Culture, language and pedagogy* (pp. 25 – 50). Norwood, NJ: Ablex.

Phillips, S. (1983). *The invisible culture: Communication and community on the Warm Springs Indian reservation*. New York: Longman.

Purves, A. (1982). Conclusion to an international perspective to the evaluation of written composition. In B.H. Choppin & T.N. Postlethwaite (Eds.), *Evaluation in education: An international review series* (Vol. 5, pp. 343 – 345.). Oxford, England: Pergamon Press.

Rose, M. (1995). *Possible lives*. Boston: Houghton Mifflin.

Santa Barbara Classroom Discourse Group. (1992a). Constructing literacy in classrooms; literate action as social accomplishment. In H. Marshall (Ed.), *Redefining student learning: Roots of educational change* (pp. 119 – 150). Norwood, NJ: Ablex.

Santa Barbara Classroom Discourse Group. (1992b). The referential and intertextual nature of classroom life. *Journal of Classroom Interaction*, *27*(2), 29 – 36.

Shavelson, R., Baxter, G.P., & Pine, J. (1992). Performance assessment: Political rhetoric and measurement reality. *Educational Researcher*, *21*(4), 22 – 27.

Short, K.G., Harste, J.C., & Burke, C. (1996). *Creating classrooms for authors and inquirers*. Portsmouth, NH: Heinemann.

Spiro, R.J., Vispoel, W.L., Schmitz, J., Samarapungavan, A., & Boerger, A. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex content domains. In B.C. Britton & S. Glynn (Eds.), *Executive control processes* (pp. 177 – 200). Hillsdale, NJ: Erlbaum.

Stake, R. (1983). The case study method in social inquiry. In G. Madaus, M. Scriven, & D. Stufflebeam (Eds.), *Evaluation models* (pp. 279 – 286). Boston: Kluwer-Nijhoff.

Tierney, R.J., Carter, M., & Desai, L. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher Gordon.

Tierney, R.J., Clark, C., Fenner, L., Wiser, B., Herter, R.J., & Simpson, C. (in press). A portfolio discussion: Assumptions, tensions and possibilities. *Reading Research Quarterly*.

Tierney, R.J., Wile, J., Moss, A.G., Reed, E.W., Ribar, J.P., & Zilversmit, A. (1993). *Portfolio evaluation as history: Evaluation of the history academy for Ohio teachers* (occasional paper). National Council of History Education, Inc.

Toulmin, S. (1958). *The uses of argument*. Cambridge, England: Cambridge University Press.

Wile, J., & Tierney, R.J. (1996). Tensions in assessment: The battle over portfolios, curriculum and control. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classrooms: Policy and practice, process and peril* (pp. 203 – 218). Hillsdale, NJ: Erlbaum.