Chapter 3


Some Consequences of Writing Assessment

Sandra Murphy

In the last several decades, the educational field at large has grown much more aware of the social aspects of learning, of literacy, and of assessment (see for example, the work of Broadfoot, 1996; Cope & Kalantzis, 1993; Gipps, 1999; Street, 1995). While learning was once viewed as the acquisition of knowledge and skills, it is now viewed as socially derived and situated, constructed and developed in interactions with others (Vygotsky, 1978, 1986). Learners, in turn, are seen as motivationally, behaviorally, and metacognitively active constructors of knowledge (American Psychological Association, 1993; National Research Council, 2000). Social perspectives have also transformed the way we think about literacy, and in particular, writing. While writing was once viewed as if it were a set of discrete skills, contemporary theories characterize it as a socially mediated process that varies in important ways across different cultures and across different contexts and purposes for writing.

In the assessment field, a good deal of work has been done to create new assessment theory and new models for assessment to accommodate these emerging social perspectives (Camp, 1993b; Frederiksen, Mislevy & Bejar, 1993; Frederiksen & Collins, 1989; Linn, Baker & Dunbar, 1990). What the new measurement theory suggests is that assessments should be congruent with the kind of learning and with the theoretical constructs they purport to measure. For the assessment field this presents a dilemma, because some traditional forms of assessment appear incompatible with new theories of learning and literacy. Yet traditional forms of assessment are well entrenched in practice, and established practices are notoriously resistant to change.

In part, this resistance stems from the competing agendas of stakeholders in the assessment process. Seen as a continuum, the stakeholders’ agendas vary from individual students using learner
centered techniques such as writing portfolios to learn and to represent themselves and their abilities to others at one end of the continuum, to government agencies looking for efficient ways to inform decisions about policy and the distribution of resources at the other (Murphy & Camp, 1996). And as White (1996) notes, teachers want assessments that acknowledge the complexity of writing and their expertise in teaching it, while testing firms and their clients want cost-effective assessments.

What has often been overlooked as a part of the “cost” of an assessment is the impact it has on students, and in particular, on the kind of education they receive. However, assessment scholars are now redefining the concept of validity to include the social consequences of assessments – their impact on curriculum and participants (Frederiksen & Collins, 1989; Linn, Baker, & Dunbar, 1990; Messick, 1989a, 1989b, 1994; Moss, 1992, 1994). And in recent years, researchers have begun to investigate this impact.

The discussion that follows reviews research on ways that statewide assessments of writing in the United States have impacted its educational environment. For the most part the scope of the discussion is limited to studies of assessments occurring in the last years of secondary school or at the point of transition from secondary school to college (or work). It is important to note at the outset that the role played by writing in assessment in this transitional secondary context in the U.S. is very different from its role in most other countries. In most national systems, extended written examinations serve as gatekeepers for university admission. The tests are designed to assess what students have learned from the secondary curriculum. In contrast, extended writing plays a marginal role at best in postsecondary admissions decisions in the U.S. To be sure, tests are used as gatekeepers to postsecondary education, but not ones that aim to measure what has been learned through extended writing. Instead, as Foster and Russell (2002) point out, “Multiple-choice tests are the primary bases for access decisions in the United States: the SAT, ACT, GRE, MedCAT, LSAT, GMAT–all the familiar exam acronyms for U.S. students–require little or no writing” (Foster & Russell, 2002, p. 330). ¹

¹ The SAT, the ACT, the Medical Colleges Admission Test, the Law School Admission Test, and the Graduate Management Admission Test.
GMAT and the MedCAT now have components that require actual writing, but those components are generally of minimal consideration in the overall performance rating of the student. They consist of brief essays designed to measure writing competence itself rather than knowledge in a particular area of study.

Other assessments at the secondary level in the US are used to monitor school progress in writing, to hold schools accountable for that progress, and to ensure that students who graduate possess minimum writing competencies. And in recent years, policymakers in the U.S. have been turning more frequently to using tests as levers for reforming curriculum, for defining, in effect, what should be taught. For instance, a survey of statewide assessment practices in 1997 showed that 46 of the 50 states had some kind of statewide writing assessment. The purpose most frequently identified by respondents (43 states) was the “improvement of instruction” (Roeber et al., 1997 cited in Mehrens, 1998). Many of these tests carry high stakes.

Stakes, as scholars have explained, are the consequences associated with test results (Heubert & Hauser, 1999; Madaus, 1988). Selection exams carry high stakes for students, because their consequences are more or less immediate and important: they determine who is qualified to study at college and university. Stakes are similarly high for students in the U.S. who are in danger of failing the high school exit exams because such exams act as gatekeepers for graduation. But because such exams set standards for minimum competency, as opposed to selection for college entrance, they may have little, if any impact on students who actually plan to attend college.² On the other hand, stakes are low for all students taking the statewide exams that are used to monitor school improvement; these

---

² For example, a recent study of high school exit exams in six states conducted by the Achieve Foundation revealed that the questions on these exams reflect “modest expectations” and test “material that most students study early in their high school years” (Achieve, 2004). States give students multiple opportunities to pass exit tests and most college-going students will have passed them long before their senior year. Unlike the Norwegian exam described in this volume, the US high school exit exams do not use writing to examine what was learned from the curriculum about subjects other than writing.
exams have no immediate consequences for the students themselves. For teachers and administrators, however, the stakes are high.

The consequences of tests used to monitor school performance affect teachers, schools and districts in very real material ways. By 1996, at least 23 states in the U.S. reported attaching explicit consequences to state test results including funding, assistance from outside consultants, warnings, and loss of accreditation (Bond, Braskamp & Roeber, 1996). More recently, the US Department of Education (2006), indicated that schools that fail to make adequate yearly progress, as determined by assessments and other factors, will undergo “corrective action”, which can include replacing school staff, instituting a new curriculum, significantly decreasing management authority in the school, extending the school year or school day, restructuring and/or takeover by the state. The high stakes for teachers and administrators attached to such tests help to ensure their impact (sometimes called backwash) on curriculum. Research suggests that tests that carry high stakes have greater impact on curriculum and instruction than tests with low stakes (Pedulla, et al., 2003). Other studies suggest that educators will sometimes respond as if stakes are high, even when the results are used only to convey information (Corbett & Wilson, 1991; Madaus, 1988).

Evidence in the research literature indicates that when high stakes are attached to tests, (or perhaps when educators perceive them as high) districts deemphasize their own objectives in lieu of those the tests reflect (Dorr-Bremme & Herman, 1986). Teachers, in turn, are influenced to “teach to the test” (Smith & Rottenberg (1991). However, whether this is seen as good or bad depends in part on the agenda of the stakeholder and in part on the nature of the construct underlying the test. Diverse approaches to assessing writing are employed in the U.S., including multiple-choice tests, impromptu writing sample tests, and portfolios. They each imply a different view of writing and how it should be taught; they each have detractors and advocates.

Multiple-choice tests of written expression

Instead of requiring actual writing, multiple-choice (sometimes called indirect) tests of writing estimate probable writing ability through observations of specific kinds of knowledge and skills associated with writing. They demand passive recognition of error and selection of best examples.
Critics of such tests have argued that they lead to negative consequences, among other things, that they narrow and fragment curriculum, discourage higher-order learning, and undermine teacher professionalism and expertise (Bracey, 1987; Corbett & Wilson, 1991; Madaus, 1988; Madaus & Kellaghan, 1993; Pearson & Valencia, 1987; Resnick & Klopfer, 1989; Shepard, 1991; Smith, 1991; Wells, 1991). In regard to writing in particular, scholars charge that multiple-choice tests fail to address the cognitive and reflective processes involved in actually producing a text (Camp, 1993b). These processes include making plans for writing, generating and developing ideas, making claims and providing evidence, organizing, establishing connections within the text, finding a tone and rhetorical stance appropriate to the audience and the subject matter, evaluating generated text, and revising, all of which require higher order skills. Ironic as it may seem, although there has been a call for rigorous standards and increased attention to writing in recent years, the most common format used in large-scale accountability systems is the multiple-choice format (Hamilton, 2003).

Yet policy makers in the U.S. value such tests as levers for reform, to move schools and teachers, so to speak, “back to basics”. Some view the impact of such tests as a positive development (see for example, J. Murphy’s (1990) review of reform efforts). As Corbett and Wilson (1991) explain, some policy makers view narrowing of the curriculum as “a focusing of the instructional program, ridding the curriculum of the distractors that have prevented schools from doing the job of providing all students with essential learning skills” (p. 133). And, in the measurement community, well-designed multiple-choice tests have been widely supported for many years because they are reliable and demonstrate concurrent validity (see for example, Godshalk et al. 1966).

**Backwash from multiple-choice tests of writing**

Research on backwash from large scale, high stakes multiple-choice tests of writing suggests that they tend to influence writing curriculum in two ways: (1) actual writing begins to disappear from the curriculum; (2) the curriculum begins to take the form of the test. For example, in case studies of two public schools in Arizona, Smith (1991) observed that teachers shifted from a writing process curriculum to “worksheets covering grammar, capitalization, punctuation, and usage” when their district’s test date neared, because those activities were better aligned with the test (p. 10). In a study
of two districts, Shepard & Dougherty (1991) reported that teachers of writing had begun to ask students to look for errors in written work instead of producing their own writing, because of the format of the writing test used in those districts. Teachers also reported increasing the amount of time they allocated to basic skills and vocabulary.

Results of a more recent investigation complement findings in these earlier studies. In an investigation of the effects of changing assessment policies in California, surveys were distributed to random samples of high school English teachers in 1988 and 2001 (Murphy, 2003). When the first survey was distributed in 1988, the state assessment plan included multiple-choice tests of written expression, but students were also required to compose a sample of writing. Using a matrix sampling plan, the state assessed a variety of genres, although individual students generated single samples of writing. When the second survey was distributed in 2001, a very different accountability system was in place. Instead of asking students to write, the assessment system at the secondary level relied entirely on multiple-choice tests to assess writing. At the time, the California Standards Tests collected actual writing samples, but only at grades 4 and 7 (California Department of Education, 2002).

Analysis of the results of the two surveys indicated that teachers who responded in 2001 when the multiple-choice test was in place spent less time teaching actual writing and more time teaching grammar and usage than teachers who responded in 1988 when actual writing samples were collected. Teachers in 2001 also reported putting more emphasis on grammar and usage and less emphasis on the features of genre in their comments on student work than teachers in 1988. The findings in this study thus support those of other studies that have investigated the consequences of assessment, namely that teachers shape instruction to match the form and content of tests.

Backwash from direct assessments of writing
In contrast to indirect, multiple-choice tests, in direct assessments of writing, students produce single (or sometimes multiple) impromptu samples of writing under timed, controlled conditions, or multiple samples of writing in a “natural” context (the classroom) supported by instruction and feedback (for example, portfolios). Such assessments treat writing as a generative process in which a variety of skills are orchestrated. In recent decades their use in accountability systems has been growing. By 1997, 37
Advocates of direct assessment argue that the best way to assess an individual’s ability to perform a task is to elicit and evaluate an actual performance of that task. Diederich (1974) put the argument this way:

People who hold the view that essays are the only valid test of writing ability are fond of using the analogy that, whenever we want to find out whether young people can swim, we have them jump into a pool and swim. (p. 1.)

Assessment scholars also defend direct assessments on the grounds that they send the message that it is important for students to learn composition skills and for teachers to teach them (see for example, White, 1995).

However, measurement experts have challenged direct forms of assessment such as essay tests because they can be unreliable (see for example, Brelang, 1983). Moreover, not all direct assessments are seen as equally valid. For example, assessment scholars have raised questions about using a single, timed, impromptu writing sample to assess writing ability. Camp, (1983), claimed that arguments about the validity of this approach rest on questionable assumptions:

1. Impromptu writing is representative of all writing
2. “Writing is writing” – the skills involved in writing are the same, whatever its purpose or circumstances.(p. 6).

Camp (1993b) also criticized timed, impromptu assessments because they are not well-aligned with contemporary, process-oriented views of effective writing instruction in the U.S. Other scholars have argued that a single sample cannot adequately represent the variety of types of writing that examinees will be expected to employ in the contexts for which most assessments are designed. For example, research has demonstrated that the writing that students do in college varies on a number of rhetorical and practical dimensions (see for example, Bridgeman, B. & Carlson, S. 1983; Hale, Taylor, Bridgman, Carson, Kroll & Kantor, 1996). Yet many tests given at the transition to college in the U.S. sample a single type of writing, one that may not align in important ways with the kinds of tasks that students in college may be asked to do.
Backwash from impromptu writing tests

Evidence on the impact of impromptu writing tests is mixed. Some research suggests that teachers are likely to increase the time students spend writing when an assessment includes this kind of writing component (Almasi et al. 1995; Koretz et al. 1996; Koretz & Hamilton, 2003; Murphy, 2003; Stecher, Barron, Kaganoff, & Goodwin, 1998). For instance, in a study of the effects of the Maryland School Performance Assessment Program (MSPAP) Almasi, et al. (1995) teachers said they altered instructional tasks “to give more writing opportunities, to emphasize personal response to reading, and to include student choice in such literacy tasks.” In a later study of MSPAP conducted by Koretz, et al. (1996), these findings were reaffirmed. Teachers reported an increase in emphasis on writing for a variety of purposes, analysis and evaluation of text, and literary comprehension along with a corresponding decrease in emphasis on spelling, punctuation, and grammar.

In contrast, other studies of the impact of such tests demonstrate a narrowing effect on the curriculum (Wallace, 2002; Scherff and Piazza, 2005), a turn toward formulaic teaching (Loofbourrow, 1994; Johnson, Smagorinsky, Thompson & Fry, 2003; Hillocks, 2002), and a negative effect on student attitudes (Ketter and Poole, 2001; Loofbourrow, 1994). In a case study of the influences of the now defunct California Assessment Program’s genre-based assessment in a junior high school, Loofbourrow found that teachers narrowed their teaching to the types, length, and forms of the test. She also found that teachers resorted to “formulaic” teaching and that students were often frustrated with the test format, either because they were prompted to write about something that wasn’t important to them or because they didn’t have enough time.

Results of Hillocks’ more recent study of the impact of statewide assessment programs on instruction in writing suggest that the trend toward formulaic teaching and writing observed in Loofbourrow’s study may be fairly widespread. After an in-depth analysis of multiple kinds of data from five states, including interviews with teachers and administrators and document reviews, Hillocks concluded that state-wide assessments will “engender formulaic writing and the thinking that goes with it” if they have the following characteristics:
… (1) prompts providing a specific topic and subject matter with no or highly limited accompanying data, (2) one
limited session for writing, and (3) criteria that call only for “developing ideas” without specifying the nature of
development much beyond a request for detail. (p. 201).

When Hillocks began his research, 37 of the 50 states in the U.S. had some sort assessment that
required actual writing. He identified twelve states that had programs with these characteristics. In
states with more effective assessment programs, criteria called for evidence, and benchmark papers
displayed their use. In states where students had more time, (e.g. Kentucky, where portfolios were
collected) they were able to find information and develop evidence for their writing. Hillock’s study
suggests that the consequences of some assessments are better than others.

Whether the impact is good or bad, however, it seems clear that the formats of writing assessment
do influence curriculum and instruction. For instance, Oneill, Murphy, Huot & Williamson (2006)
found that test formats in three states influenced what teachers taught and how. The assessment
formats in the three states were very different. The “writing” test in California was a multiple-choice
test. In Georgia, the test asked for a single sample of persuasive writing produced under controlled,
timed conditions. In Kentucky, students were required to produce portfolios of writing under normal
classroom conditions with a focus on “writing for authentic audiences” and “situations” (Kentucky
Department of Education, 1999).

Questions on the surveys distributed in the three states asked teachers about the influence of tests
on curriculum and about their teaching practices. The results indicated significant differences across
the three states in the types of writing that were assigned most frequently, the length of assignments,
the number of drafts and the amount of time allowed for the writing assignments. In California, where
there was no actual writing sample collected, 72 % of the teachers assigned response to literature, a
form of writing that aligned closely with the typical content of their English courses, most frequently.
In Kentucky 52 % of the teachers assigned response to literature most frequently, but 48 % of them
assigned a variety of other kinds of writing instead, including types of writing such as short story and
autobiographical narrative. In Kentucky, the assessment system called upon students to choose pieces
that addressed a variety of audiences for a variety of purposes, using a variety of forms. In Georgia
48% of the teachers assigned response to literature most frequently, but 52% of them assigned other types of writing and those types aligned in important ways with the type of writing on the state test. In California and Kentucky, where writing was collected under “natural” classroom conditions, teachers were more likely to allow three or more days for writing, to require three or more drafts, and to require assignments of greater length. Teachers in Georgia, on the other hand, were more likely to assign single-draft, short pieces of writing and to require that the writing be turned in the same day than teachers in California and Kentucky, a finding that was consistent with the 90 minutes allowed for writing on the Georgia state test. The results of the survey thus support the idea that test formats influence not only what types of writing are taught, but how. Evidence from other research suggests that teachers not only focus on the types of writing covered in a test, they shift their instruction and evaluation strategies to match the rubrics used to score the assessments (Mabry, 1999).

**Backwash from portfolios**

As an assessment method, portfolios appear to address many of the concerns held by assessment scholars in the U.S. about the validity of using timed, impromptu, single sample assessments to assess writing ability. Because portfolio samples are collected under more natural and authentic conditions and embedded in instruction, expectations can be made transparent. Further, when a portfolio approach is used, writing can be treated as a recursive process and revisited for revision. Moreover, evidence from research shows that time and instructional support for writing give students a better chance to do their best (Hillocks, 2002; Herman et al. 1993). As noted above, Hillock’s analysis suggests that students in Kentucky, where portfolios were assessed instead of impromptu samples of writing, students had time to find information and develop evidence for their writing. Herman et al. (1993) found that raters’ scores for portfolio collections of classroom work were higher than those for a standard writing assessment in which students were given 30 to 40 minutes to write. Portfolios also offer opportunities to broaden the assessment construct by sampling a range of genres and to engage students more directly in the assessment process in ways that give them responsibility for evaluating their own learning (Camp, 1992, 1993a).
However, the emphasis here is on the word “can.” Because they reflect curricula and the web of beliefs, goals, and assumptions that underpin education, portfolio programs vary widely in ways that are theoretically and educationally important (Murphy & Underwood, 2000). In some assessment programs, portfolio contents are highly standardized. In others, students are encouraged to choose what the portfolios will contain. In some programs, students submit only final, polished versions of work; in others they submit drafts, notes, and other evidence of the processes they engaged in as they produced their work. In some programs, students are asked to reflect on their work, their learning, and/or their processes, in others, not. All of these basic differences in assessment design reflect at some level different views about the construct of writing that the assessment purports to measure and the role of the student in the process. Not surprisingly, like other approaches to assessment, the results of research on the consequences of portfolio assessment programs are mixed.

Some studies of large-scale portfolio assessment programs have demonstrated positive effects on the educational environment. Nearly three-fourth of the principals interviewed in a study of Vermont’s portfolio assessment program reported that the portfolio program produced positive changes, including “an increased emphasis on higher order thinking skills,” “lessened reliance on textbooks and worksheets; an increase in writing overall and more integration of writing with other subjects; more work in cooperative groups … ” (Koretz, Stecher, Klein & McCaffrey, 1994, p. 31).

Evidence in the literature also suggests that participation in scoring sessions for this kind of curriculum-embedded assessment contributes to teachers’ knowledge and expertise and to curricular reform (Gearhart & Wolf, 1994; Sheingold, Heller & Paulukonis, 1995; Storms, Sheingold, Nunez & Heller, 1998). For example, Storms et al. (1998) found that teachers learned about the qualities of student work from the conversations that occurred during scoring sessions. Teachers surveyed in a study by Sheingold, Heller, and Paulukonis (1994) reported substantial changes in their teaching and assessment approaches. In particular, they indicated changes in the sources of evidence they used for assessing student performance, in their expectations for students’ responsibility for their own learning and assessment, in their goals for instruction, and in their use of explicit performance criteria to evaluate student work. Finally, using and triangulating a number of data sources, Gearhart and Wolf
(1994) found that teachers increased their understanding of the narrative genre and their capacity to provide focused, genre-based comments about children’s narratives.

Portfolio assessments, however, have not exerted uniformly positive effects, in part because the design of portfolio assessment systems can work against instructional reform goals. One study, in which a highly standardized portfolio was employed, found that because so many different types of writing were required, students had few opportunities to practice and refine any one type, had few decisions to make about the relative quality of different pieces of their work, and were thus unable to reflect on individual progress and goals (Murphy, Bergamini, and Rooney, 1997). Similarly, while Kentucky with its statewide portfolio system fared much better overall than other states in Hillock’s (2002) study of the impact of statewide assessments on writing instruction, other research suggests that the system has had negative consequences. Callahan (1999) found that some high school English teachers in Kentucky see the portfolios “primarily as a stressful administrative task” … “imposed from outside, introduced as a high-stakes accountability task, and embedded in a massive top down reform effort.” In her view the pressure of the assessment situation also encouraged “a form of dishonesty among both teachers and students” when some portfolios were accepted despite questions about the origin of some of the texts they contained (pp. 34–35). Spaulding and Cummins (1998) found that many students at the University of Kentucky who had completed the portfolios said that compiling them was not a useful activity (p. 191). Concerns have also been raised by measurement experts about the cost and the difficulties of obtaining reliable scores for portfolios (Nystrand, M. Cohen, A. & Dowling, N. (1993), although some assessment programs have overcome those obstacles (LeMahieu, Eresh & Wallace, 1992; Herman, et al., 1993).

Conclusion
Although direct cause and effect relationships cannot be demonstrated, taken together the findings of the studies reviewed here suggest that the content and format of statewide accountability writing tests influence curriculum and instruction. When policy-makers make decisions about the design of such assessments, they will necessarily be making choices about what kinds of writing or writing skills will be tested, and in the process they will be privileging some curricular content represented in the
assessment and ignoring other content that is not. They will also be privileging particular instructional approaches. In the present high stakes environment in the U.S., it is likely that their choices will prompt changes in teaching.

Given the powerful interplay between curriculum and assessment, it is all the more important to design assessment systems that enhance learning. Although there are no guarantees that authentic assessments will do so (Mehrens, 1998), they are certainly more likely to enhance learning than assessments that are inauthentic. Authenticity refers to “the degree of correspondence of the characteristics of a given language test task to the features of a target use … task” (Bachman & Palmer, p. 23). For writing assessment this means the test writing tasks should represent the types of writing that examinees will be expected to employ in the context for which the assessment is designed. Unlike multiple-choice tests, authentic assessments engage students as active constructors of meaning. Thus they are in line with current views of learning. When assessment is authentic, and when the activities involved in building and conducting the assessment also contribute to the professional development of teachers, some would argue, the assessment itself can become an effective means for promoting the kinds of learning it addresses (Eresh, 1990; Camp, 1992b, 1993a; LeMahieu, Gitomer & Eresh, 1995; Moss, 1994).

It is also clear there is a need for a more integrated, systemic approach to assessment, one that takes into account the impact that assessment necessarily has on curriculum. When assessments are systemically valid, they “induce curricular and instructional changes in education systems (and learning strategy changes in students) that foster the development of the cognitive traits that the tests are designed to measure” (Frederiksen & Collins, 1989, p. 27). Arguments against authentic assessment often rest on the grounds of technical adequacy and cost. An important challenge for measurement experts and policy makers in the future, then, one that is especially relevant to the consequential validity of assessments, will be to balance the demands of technical adequacy and cost with the development and validation of assessments that enhance learning and the educational environment. As Frederiksen and Collins (1989) point out, “the efficiency in current testing practices is greatly outweighed by the cost of using a system that has low systemic validity—one that has a negative impact on learning and teaching” (p. 32).
However, determining whether the impact of an assessment is negative or positive is a problematic issue. Assessments of writing necessarily embody different priorities and views of learning because they are typically structured around particular content, a particular taxonomy of discourse, or a particular curriculum for writing. As the research reviewed here reveals, whether consequences of an assessment are viewed as good or bad depends on the values and beliefs of the stakeholder. While some curriculum and assessment experts value new modes of assessment because they are in line with current views on learning and because they assess higher-order skills as opposed to rote reproduction of knowledge, not all stakeholders hold the same beliefs and values. Herein lies an intractable problem for large-scale assessments. Although assessments can be designed to further particular instructional goals, the goals themselves remain matters of public debate. Another intractable problem for the developer of a large-scale assessment, then, is how to balance the competing curricular agendas of its stakeholders.

To address this problem, small-scale alternatives to large-scale writing assessments have been proposed by some assessment experts (see for example, Broad, 2003; Huot, 2002). For instance, Huot (2002) believes assessments should be site-based, locally controlled, context-sensitive, rhetorically based, and accessible. A “site-based” assessment, Huot explains, “is developed in response to a specific need that occurs at a specific site,” and that is based on the “resources and concerns” of its constituents. In a “locally controlled” assessment, the individual institution would be “responsible for managing, revising, updating and validating the assessment procedures, which should in turn be carefully reviewed … to safeguard the concerns of all those affected by the assessment process.” In a “context-sensitive” assessment, the procedures would “honor the instructional goals and objectives as well as the cultural and social environment of the institution or agency and its students, teachers and other stakeholders.” In a “rhetorically-based” assessment, “All writing assignments, scoring criteria, writing environments and reading procedures” would “adhere to recognizable and supportable rhetorical principles integral to the thoughtful expression and reflective interpretation of texts.” In an “accessible” assessment, all “procedures and rationales for the creation of writing assignments, scoring criteria and reading procedures, as well as samples of student work and rater judgment” would “be available to those whose work is being evaluated” (p. 105).
Were such principles to be enacted, the resulting assessments would be much more likely to accommodate the interests and curricular goals of their particular stakeholders. But Huot speaks from the perspective of a college writing program administrator and teacher. Given the widespread use of large-scale assessments to monitor educational programs in the U.S., and their popularity with policy-makers and the public at large, it is unlikely that locally-controlled, site-based alternatives will ever be considered outside of the college environment. Indeed, policy-makers want large-scale measures that allow them to compare individual sites to one another, using, so to speak, the same yardstick. For public education then, at least at the K-12 level, the dilemma remains: how to balance the need for efficient, cost-effective assessments with the need for assessments that are aligned with contemporary views on literacy and that enhance the educational environment.

References


