



The Politics of High-Stakes Writing Assessment in Massachusetts

Why Inventing a Better Assessment Model is Not Enough

DAN FRAIZER

Springfield College

What happens when government officials conspire with a national testing company to control literacy standards for teacher preparation students on a statewide level? This essay documents the politics of the Massachusetts teacher test story, focusing on the flawed process that led to a writing test that excluded the participation and negotiation of stakeholders. I argue that as a discipline, we need to learn to play politics better, faster, and with a strong disciplinary commitment to promoting assessment models that are fairly negotiated. Writing professionals should organize in order to participate directly in good faith discussions with powerful interests so as to promote locally developed and decentralized assessment models.

The emerging new world order of writing assessment as defined by state and corporate authorities challenges all writing teachers to both understand the state-mandated tests created by testing corporations and to openly discuss how we should build a movement to challenge top-down assessment in all its forms. Writing assessment can no longer be something we pretend to do on our own within the confines of academic institutions. As Huot and Williamson (1997) argue, power is an element of the assessment process that cannot be ignored, even though we in academic institutions have often sidestepped questions of who controls writing assessment and curriculum in the schools. In the meantime, state

Dan Fraizer teaches first year writing courses, directs the Writing Center, and supports the efforts of teacher preparation students to pass the state-mandated teacher tests at Springfield College in Western Massachusetts. His essays and research have appeared as articles in *The Writing Instructor*, *Teaching English in the Two Year College*, and *Michigan Journal of Community Service Learning*, and as book chapters in *The Literacy Connection* and *The Politics of Writing in Two Year Colleges*.

Direct all correspondence to: Dan Fraizer, Department of Humanities, Springfield College, Springfield, MA 01109-3797, dfraizer@spfldcol.edu

departments of education exercise their power to control assessment with increasingly popular efforts that betray their ignorance of assessment issues and priorities. We must learn to play politics better, faster, and with a strong disciplinary commitment to promoting assessment models that are fair and negotiated.

In this article, I will first review some of the many controversies surrounding high-stakes standardized testing in general and writing tests in particular. I then discuss relevant assessment and literacy scholarship in order to reinforce arguments by others that assessment tools should be developed in an atmosphere of local, contextualized negotiation, not the inherent conflict associated with high-stakes tests. I also narrate the politics of the Massachusetts teachers' test story as an illustration of what happens when negotiation and mutual respect is absent because government officials conspire with a test maker to control literacy standards, exclude the participation of stakeholders, and arrogantly defy criticism.

Writing assessment has always been linked to politics, especially since the 1970s when composition as a discipline defined itself against the authoritative current-traditional rhetorical model of language instruction. The struggle has always been about who controls the standards and how they are created. Writing assessment has been described as a form of social action that can either constrain or empower. When assessment empowers students and teachers, it is because various stakeholders have made room for reflection in the process of creating assessment models, as "a means to foster critical inquiry into literacy" (Huot & Schendel, 2001, p. 51).

My purpose in this article is first to raise awareness about the deeply flawed process that led to the creation of this test, a process mired in assumptions that composition as a discipline has always intended to challenge. I also argue that writing professionals of all kinds should participate directly in good faith negotiations with powerful interests in order to create writing assessment tools that are consistent with our values as a profession, and that we must work to elect powerful stakeholders in government who represent our interests because we cannot always assume these powerful interests will be willing to work with us in good faith. We must also do a much better job of educating the public about the risks associated with high-stakes literacy tests. Generally, I hope to persuade readers that national organizations like College Composition and Communication, the National Council of Teachers of English, and their regional affiliates should be doing as much as possible to organize us so that we can support state officials who will do the right thing, and when necessary, challenge state officials and testing companies when they assess writing in narrowly defined, discriminatory ways that privilege profit over students' interests and promote a conservative, hegemonic political agenda.

Standardized Testing in the United States

The history of the standardized testing industry, as chronicled by Lehman (1999) in *The Big Test* does not inspire confidence. Carl Brigham, a psychology professor at Princeton in the 1920s, was responsible for the first U.S. Army IQ tests, which eventually became the first Scholastic Achievement Tests (SATs). By the 1930s, Brigham recanted much of what he had written about the connection between SAT scores and intelligence, and wrote strongly in favor of

the SATs as predictors of academic performance only. He also questioned one organization's ability to develop, research, and market such tests. But his work was eventually appropriated by those who were more interested in genetic predisposition and the use of tests to label test takers' innate mental abilities. The belief that the SATs could be used as IQ tests was promoted by Henry Chauncey, who created the Educational Testing Service (ETS) at Princeton. These views became so entrenched that they were not successfully challenged until Ralph Nader's inquiry in the 1970s showed scores did not reflect a genetically fixed standard but could change with coaching (Lehman, 1999).

The story of ETS illustrates the extent to which a company financially exploited the public's faith in scientific objectivity and higher education's need to legitimize its system of meritocracy. Founded as a nonprofit corporation in 1947 with a meager endowment from the Carnegie Foundation, by 1998 ETS had 100 testing programs, a budget of \$460 million, and 2,400 employees across the country. ETS has in recent years successfully weathered further legitimacy challenges. Although its tax-exempt status was challenged in 1997, ETS has signed multi-million dollar contracts with Sylvan Learning Systems to electronically deliver ETS admissions tests (GRE and GMAT) both in the United States and abroad (Sacks, 1999). Might the ETS of today, which considers itself to be a global corporate enterprise, be philosophically in conflict with the ETS that considers itself to be an organization serving the public?

What standardized tests often seem to measure best is socioeconomic status. The College Board, which "sponsors" the SAT, has argued that the SAT is not designed to predict success in college or the workplace. Instead, questions that require abstract problem-solving skills are made to represent general intelligence. It's no surprise that success on standardized tests is most strongly correlated with the values and learning of those in higher socioeconomic classes, so much so that "one could make a good guess about a child's standardized test scores by simply looking at how many degrees her parents have and what kind of car they drive" (Sacks, 1999, p. 8). How reflective of status are standardized tests? When James Gee asked a group of honors students to answer multiple-choice questions from the reading portion of the SATs without ever reading the corresponding reading passage, about 80% of the 100 students answered all of the questions correctly. Gee noted that "mastery of an [elite] cultural model was what was actually being tested" (cited in Luna, Solsken, and Kutz, 2000, p. 284). Elites in the United States understand how important standardized tests are to maintaining their privileges. Recent reports show that the group most likely to claim to have a learning disability in order to win extra time to complete the SATs is rich, White teenage boys, especially from Connecticut private preparatory schools (Weiss, 2000).

Writing Assessment and Literacy Research versus the Market

Some writing professionals have been influenced in recent years by what has been called the "New Literacy Studies" (Gee, 1990; Street, 1984, 1998). This work has emphasized the primacy of local contexts in defining literacy practices and assessment. These theorists argue that literacy should be understood as a

social practice, and that assessments cannot be generalized across diverse, cross-cultural settings. To support these arguments, case studies illustrate the diverse ways reading and writing are used and evaluated. Whether texts are meaningful or not depends on the social practices within which they are embedded. Writing theorists have also emphasized the importance of holistic models of assessment best negotiated in local contexts (Huot, 1996; Huot & Williamson, 1993; White, 1998).

Sociocultural literacy critics challenge the legitimacy of an “autonomous” literacy, along with some of the reliability concerns of many assessment experts, when they argue that fixed standards do not take into consideration locally defined goals and outcomes. The autonomous and narrowly defined literacy models sometimes valued by government assume one standard can be applied across diverse and expanding literacy situations in the workplace (Street, 1998) and have challenged researchers to try to create more reliable, valid, and fair forms of standardized assessment (Huot & Williamson, 1993), or to propose less standardized forms altogether (Moss, 1994). Critics of “indirect assessment,” or machine-scored tests, challenged their claims of objectivity and value-neutral status by understanding “both reading and writing as the construction of meaning” (Huot & Williamson, 1993, p. 3). A major response to this criticism has been holistic writing sample evaluation done collectively with an emphasis on overall meaning. But holistic evaluation has been criticized for slighting both context and process in the interest of efficiency and reliability (Huot, 1996), and critics of all forms of standardized testing, including holistic models, now ask questions such as, “Who decides if you have been successful and on what basis?” “Who decides what is worth testing?” and “What are the consequences of using high-stakes tests that are informed by an autonomous perspective on literacy?” (Luna et al., 2000, p. 285). The origins of such challenges to a fixed standard of literacy assessment can be found in the work of writing scholars such as Mina Shaughnessy (1977), who was one of the first to argue that impromptu writing does not measure a writer’s true ability. Both high-stakes and impromptu writing assessments have also been criticized on the basis of race and sex discrimination (Camp, 1996; Holdstein, 1996). Although the implied purpose of all tests is to measure in order to sort students by ability, critics argue that standardized tests do not sort students by ability alone, but by other embedded, often unseen and unspoken criteria that result in the privileging of certain groups of students over others. Huot and Williamson (1997) have argued that assessment is political because of the concentration of power in the hands of those who decide how assessment will be standardized, often times “based on schemes that are so pared down by standardization that they produce information that has little meaning and importance for local contexts” (p. 53).

Some of the most invisible and overlooked criteria of autonomous literacy models are those important to the companies that create the tests. White (1996) notes the following assessment criteria are most valuable to testing companies:

1. The assessment produces scores quickly and cheaply.
2. The assessment reduces the complexity of writing and the teaching of writing but allows the data collected to imply complex measurement.

3. The assessment weighs heavily surface features of writing and dialect features of edited American English.
4. The assessment leads to the sorting of students according to existing social patterns.
5. The meaning of assessment depends heavily on statistical explanations of sufficient complexity to invite misuse of scores.

In other words, the criteria most important to testing companies are criteria that oversimplify the complexity of writing assessment in order to enhance the companies' credibility, lower production and administrative costs, and legitimize well-established sorting patterns. Testing companies do this for the money. Sacks (1999) cites the following: "By 1997, standardized achievement test sales in the K-12 market alone had reached more than \$191 million, and that was up more than 21% in real, inflation-related dollars in five years" (p. 222). These profits have been fueled by the mandating of tests by federal and state governments. The proposals put forward by the Bush administration to mandate testing in all the states through the "no child left behind" program would create even greater profits for these companies. State government mandates alone have been a bonanza for testing companies. Harcourt Educational Measurement, a division of Harcourt Brace Jovanovich, is one of the leaders in sales of testing products in the United States. In the first 9 months of 2000, their profits increased by 38%, and theirs is only one story of explosive growth (Pierce, 2003). In light of such profits, is it possible that the most difficult questions now being asked about writing assessment are simply ignored by those responsible for creating high-stakes tests?

Testing and State Government Involvement

The relationship between state governments and institutions preparing teachers was not always as confrontational as it now appears to be. Initially informal, it was only after World War II that more formal agreements developed between states and colleges concerning the certification of teachers. States began to approve teacher education programs, thereby "entitling" teachers to a teaching license. Then, beginning in the 1980s, and fueled by various educational "crises," many states no longer trusted teacher preparation programs to ensure "quality" teachers in the classroom. By the late 1990s, more than 40 states used teacher tests in addition to program "entitlement" for certification (cited in Melnick & Pullin, 2000).

These tests have been categorized as "high stakes" because of their ability to prevent students from becoming certified as teachers, thereby negating all other tests a student might take in their college courses. A central assumption of high-stakes tests is that their presence will pressure teachers to enforce a minimal standard of proficiency (and hence "quality") in specified areas of student ability. High-quality teachers supposedly ensure that children pass high-stakes tests. The public debate around the country has often seemed so focused on "quality" that some states, such as Arizona and Colorado, have criticized their own teacher tests as "too easy" and "meaningless" because the majority of test takers pass the tests (Scanlon, 1998; Schultze, 1992).

But the content of these tests is rarely examined or understood in terms of conflicting ideological perspectives represented in the test questions or whether the tests measure a candidate's ability to teach. Pence (1998) argued that standards setting is one element of educational reform, but to ground standards in positivistic models or the rhetoric of crisis and accountability will not improve the quality of teaching and ignores the ongoing disciplinary construction of such standards that must be a part of authentic reform.

Why would there be a perceived need to certify the literacy (as opposed to the teaching ability) of teachers? First, of course, is the public perception that literacy is at the heart of what goes on in schools. Test proponents describe the tests as a screening device: "These tests don't show that anybody will be a good teacher," said one board of education member. "Failing them shows only that you are, thus far, incompetent in some respect that is essential to good teaching" (Bradley, 1998, p. 34). Test opponents take issue with the tests themselves, which they argue are not valid, reliable, or even legal means of measuring skills necessary for job performance (Melnick & Pullin, 1999). Neither side argues that teachers should be illiterate, but proponents and opponents are divided over who develops literacy standards and how literacy should be measured. The conflict can be intensified by a publicly constructed sense of "crisis."

Fox (1999) made the distinction between standards a teacher holds "as provisional goals for his or her students, which develop out of and are modified by interactions in context, and bureaucratic standards that almost always emerge from a political context of crisis" (p. 10). This crisis mentality has led critics of high-stakes tests in general and teacher tests in particular to point out two common problems with hastily constructed bureaucratic standards: their tendency to restrict access of members of minority groups in disproportionately large numbers (Hood & Parker 1991), and the undemocratic process by which they are created, especially who makes decisions about test content and form. Alternatives do exist. Delandshere and Petrosky's (1998) efforts to create a national teacher assessment procedure that supports local certification decisions is one such model, but local control has thus far meant a race to produce the "toughest" and most politicized systems of certification usually based on quantification systems that are themselves suspect because they reduce the validity of complex performances (Delandshere & Petrosky, 1999).

States now wield great power in their ability to control who is considered literate. In Texas, a high-stakes accountability law requires teacher preparation programs to raise their pass rates on the Texas teacher tests or be shut down. But Texas colleges that serve primarily Hispanic or African-American students have struggled to meet state of Texas mandates, making it less likely that minority teachers will enter the classroom (Blair, 2001). The same is true in California, where a long-running lawsuit claims that California's teacher testing program is biased against minority candidates because it does not measure a prospective teacher's skill in the classroom and has a negative impact on minority members (Hoff, 2000).

State government and educators can be divided over who will define and control test standards. In the state of Washington, where a "Professional Educator Standards Board" advises the state board of education on the content of its newly approved teacher test, 7 of 19 members of the board are public school teachers, but

all members are appointed by the governor. The names of all prospective members of the standards board are submitted by Republican and Democratic caucuses of the state House and Senate, exemplifying an explicitly political, not “scientific” or “objective,” structure for creating test standards (Bradley, 2000).

Examples of direct collaboration between state governments and educators in the creation of state-mandated tests do exist. Hillocks (2002) documented state-mandated writing assessments of K-12 students done in Illinois, Texas, Kentucky, and Oregon. Illinois and Texas enlist the involvement of selected teachers from around the state who serve on advisory committees, but in each state, it is the state department of education and state legislatures that have the final say about how these tests will look. Authentic collaboration seems to be a rarity, as the advice of these committees may be ignored.

Brinkley and Sudol (1999) described the efforts of the Michigan Council of Teachers of English (MCTE) to collaborate on a high-stakes high school literacy test in Michigan. The creators of the tests aimed to show that “defining and measuring literacy within the context of standardization needs to be a negotiated process requiring deliberation, flexibility, and patience on the part of numerous stakeholders” (p. 46). In the end, MCTE was not entirely pleased with the test that was eventually created, nor with the power Michigan politicians acquired to use the tests to reduce school funding, promote charter schools, attack public school teachers’ credibility, and reward those students with high scores. But a wide range of vested interests were represented. K-12 classroom teachers and administrators, teacher educators and rhetoric and composition specialists, state department of education testing and measurement staff, staff from a national testing company, legal consultants, state-level business and professional organizations, and members of the Michigan State Board of Education and Michigan State Legislature all collaborated to create a high-stakes high school graduation test (Brinkley & Sudol, 1999, p. 45). Addressing the concerns of many participants meant validating the functional and quantifiable literacy dictates of the Michigan Department of Education and testing company representatives, as well as the critical literacy values of teachers, administrators, and researchers, which included “reflecting on one’s own writing processes, using writing to communicate about a subject other than English language arts, and writing to generate and convey original ideas that are thoughtfully developed and polished in response to a specific writing task” (p. 45).

To some, the Michigan model of negotiating priorities was productive even though MCTE did not get all that it wanted. To others, collaboration simply led to a different set of problems, mostly related to losing local, contextualized evaluation settings. For example, how could students anticipate whether the reflections they committed to paper would be considered acceptable when they didn’t know who would be reading their writing? Discussion of prewriting responses was meant to address the problem of isolation inherent to most high-stakes testing situations by attempting to replicate the consultation and interaction with texts and with others that occurs in real-life writing situations. But, in the end, Michigan students continued to write in isolation from one another, not knowing who would be reading their writing or what readers’ views about the topic would be. Since its implementation, some elements of the Michigan test, such as the reflection component, have

been challenged by the state, and when organized opposition to the test has arisen, attempts have been made to diffuse criticism by offering financial awards to those with high scores. Does collaboration lead to a better test than would have been created with no collaboration? The answer is probably “it depends on what you mean by collaboration.” The greatest risk is that testing companies and state governments will misuse or ignore the input from advisory panels or oversight committees and simply “wear down” educators through the creation of a pretense of involvement, with no guarantee that they will actually respond meaningfully to their input.

The Massachusetts Teacher Tests

The struggle over teacher certification testing and assessing “minimal literacy” in Massachusetts since the late 1990s has pitted many educators, including writing program administrators and teachers, against state politicians and bureaucrats. In Massachusetts, politicians, State Board of Education officials, and a testing company created certification requirement tests for prospective school teachers in a top-down effort to attempt to improve the quality of the schools in the state. The Massachusetts Tests for Educator Licensure (MTEL), formerly the Massachusetts Educator Certification Tests (MECT), were first administered in Massachusetts in April 1998, and include a subject area test and a literacy test. The literacy test is divided into two major parts—reading and writing. About 80% of the “Reading Subtest” is comprised of multiple-choice questions intended to measure a test-taker’s understanding of brief reading passages. Another 20% is comprised of “fill-in-the-blank” word definitions. Test takers are required in this section to define words such as “democracy” as though they were writing the dictionary definition of the word. The “Writing Subtest,” the more controversial of the tests, is comprised of four sections. In one section, test-takers are required to read a 750-word passage and summarize its main ideas and essential points in 250 words using edited American English. In a second section, test-takers are required to compose an “extended written passage on a specified topic of general interest” that is “unified, well organized and developed, and written in edited American English” (Massachusetts Department of Education, 1998, p. 5). In a third section called “Grammar and Usage,” test-takers must identify “grammatical, usage, and structural errors” in written passages through multiple-choice responses, identify grammatical and usage errors in sentences and rewrite the sentences in edited American English, and again write “word definitions,” this time defining “grammatical terms” by responding to statements such as “Define a preposition.” In the final and most controversial section, entitled “Written Mechanics,” test-takers must listen to an audiotaped reading of a 150- to 200-word passage and “write the passage down word for word, using edited American English and correct spelling, capitalization, and punctuation.”

The Massachusetts writing test is notable first in terms of its sheer bulk. Few, if any, literacy tests for teachers in other states require so much effort to complete. By comparison, the Praxis Series tests, developed by Educational Testing Service, are used to evaluate teaching candidates in at least 35 states around the country. The

Praxis Writing test, however, is limited to a 30-minute multiple-choice test (error identification) and a 30-minute essay section. Students are officially allowed at least 2 hours to complete the writing portion of the MTEL, but many students are advised to intentionally fail other parts of the test on their first attempt at taking the MTEL so they will have more time to work on the writing test. Students are allowed to retake the tests as many times as they like in order to pass, and once they have passed one section, such as writing, they no longer have to retake it. Each time they retake the test, they are required to pay the full registration cost.

How did teaching candidates from Massachusetts end up facing such difficult writing assessment demands? Hillocks (2002) pointed out that “politicians who wish to demonstrate their capacities as educational leaders will not want a test that is too demanding’ because they won’t be providing the nontesting resources necessary for teachers and students to meet the demands of a test that is too difficult (p. 70). But what if these authorities don’t care if some students don’t meet the demands of the test? Whose interests are served by creating tests that limit the number of children who can graduate from high school or limit the number of teachers in teacher preparation programs from entering the profession? Like other prospective teachers around the country, Massachusetts teaching candidates are attempting to enter the teaching profession during a time when state and federal governments are more involved than ever in education initiatives. But what are the goals of these initiatives? If the intent is to improve educational opportunities, why is there such animosity between educators and government officials? These questions can first be addressed through critical assessments of the testing industry and an awareness of critical literacy theory.

How It Happened in Massachusetts Is How It Could Happen Elsewhere

The Massachusetts teacher tests were born in 1985, when the Massachusetts legislature passed a law requiring teacher candidates to pass an exam that measured both their expertise in a subject area and their “communication and language skill.” The reality of Massachusetts politics was such that nothing was done to create this test until June 1993, when the legislature enacted the “Education Reform Act,” which required candidates to specifically pass “a writing and subject test” that would qualify them for certification. Writing ability and the student’s area of expertise were clearly high priorities in the minds of these legislators. In 1996, the State Board of Education requested applications from teacher certification test contractors while promising that the public would be informed and consulted. In the space of a year, advisory committees reviewed test material, pilot tests were conducted, and test materials were submitted to the Department of Education. The commissioner of education at the time, Robert Antonucci, selected *National Evaluation Systems* (NES) to develop and administer the tests, with the caveat that the tests would be “rigorous and of high quality, based on college-level content” (Haney, Fowler, & Wheelock, 1999, p. 51).

It is unclear whether Antonucci was aware of the problems NES had been facing. The teacher test created by NES for Alabama in 1989 was eliminated after a

lawsuit by African Americans challenged its validity and arbitrary cut-off scores. In New York, another NES exam was challenged in court as biased and unrelated to job performance, since the Americans with Disabilities Act of 1992 requires that literacy tests be nondiscriminatory and related to specified job requirements (Schaeffer, 1998, par. 9).

Registration for the test began in January 1998, with test-takers first assured their scores would not count toward certification, but then told in March the scores from tests taken in April would count. Very little information about the test was provided. NES did not want the initial scores to count because they wanted to use this time period to validate the test. But by then the new test was in the news and had become a political issue debated by gubernatorial candidates. With the public supporting “get tough” policies on teacher “quality,” the issue of test validity was neglected in the media.

The State Board of Education, probably under political pressure from the governor, saw no reason to wait to count the initial rounds of test scores (April and June), and also recognized no conflict of interest in their own meddling in the process of test preparation or content. A study guide for the test, including a promise of future sample questions and guides, was finally issued in March 1998, but then withdrawn because of a last minute change in the Written Mechanics Exercise (cited earlier) to require test-takers to copy correctly a section from the *Federalist Papers*. NES would take another year (July 1999) before submitting its own 1,700-page report on the reliability and validity of the test. That report was in turn criticized by testing experts as inconsistent with educational and psychological testing standards (Ludlow, 2001) and as a conflict of interest since the test was validated by the company that created it (Associated Press, 1999). Eventually the Department of Education appointed a technical advisory committee to report on the test. Although the committee made at least 13 recommendations to the state, including a mandate for closer cooperation with college educators, almost none of them was implemented. However, the committee did report that the testing program was “a strong, sustainable, psychometrically sound and essential component of [the state’s education] reform effort” (Driscoll, 2002), thereby validating the test and defusing any effort to revise or dismantle it.

Further examples of state meddling in the creation of the test abound. By April 1998, NES had developed scoring panels to recommend cut-off scores (pass/fail) for each test to the State Board of Education. In June, the Board of Education voted to set the cut-off score at one standard error of measurement below the scores recommended by the panels, probably because of the low pass rate during April. But by July, under pressure from the governor, the Board met, rescinded its earlier decision, and set the cut-off scores at the level initially recommended by the panels.

Another example of state involvement in the creation of the test is more directly related to writing. We know that the dictation section of the Writing Test was suggested by State Board of Education members John Silber and Edwin Dellatre, and that during the first administration of the test, both men convinced the test-makers that test-takers be required to correctly copy a paragraph from a “great” writer. The test-makers complied by choosing a section from the *Federalist Papers*, written 200 years ago when ideas about “correctness” were very different than they

are now. To my knowledge, no other teacher test in the nation requires students to perform this odd exercise, which seems unlikely to be endorsed by most researchers in writing or literacy.

Perhaps meaningful cooperation among the various vested interests would have reduced the likelihood that the MTEL writing test would end up looking the way it did. But government and industry conspired to keep most educator voices silent. NES and the Board of Education did field-test questions by soliciting responses from some 300 unnamed “educators” across the state, but we have no idea whether their input changed the tests at all. Secrecy surrounding the contents of the Massachusetts teacher test was a priority from the beginning, with minimal study materials and no actual practice tests provided by NES (because of the cost, they said). The effort by NES to suppress information about the MTEL supposedly enabled NES to compete with other testing companies by controlling access to their tests. But some of the data about the test that should have been known was either missing or problematic. For example, NES revealed that no non-White candidates passed all sections of the test on its first administration, and that it stopped collecting data about race after that (Melnick & Pullin, 1999), thus following in the footsteps of ETS to suppress racial problems with their tests (Lehman, 1999). NES also admitted it did not have documentation requested by administrators concerning test development, test bias, or the background of the people who had created the test (Associated Press, 1998). Frank Haydu, interim commissioner of education, resigned in protest because of his concern that the state would be sued by someone failing the test due to “racial and ethnic implications” (Appleton, 1998, p. A1).

State politicians also participated in keeping test criticism under control. When David Driscoll, the third commissioner of education in 3 years, criticized a detailed, 78-page report from local experts in educational assessment (the Ad Hoc Committee to Test the Teacher Test) by accusing the authors of having the “wrong focus” and “that the real story was the lack of skills among many of the candidates,” *Phi Delta Kappan* characterized Driscoll as “a man who clearly flunked Educational Measurement 101” (Gough, 1999). Belligerence in the face of test criticism was epitomized during a panel discussion at a western Massachusetts community college, where Driscoll and others took questions from parents, teachers, and administrators, while the media looked on. When a test critic charged that the decisions were made about the high-stakes tests from the top down, a local state senator responded with this accusation: “We need civil discussions. That was a veiled threat. The majority of states in America are doing this. Can so many other people be wrong?” (Tantraphol, 2000, p. A8).

When questions about the tests were occasionally cited in the media by testing experts, teacher education administrators, or students taking the test, state officials were usually quoted as denying that the topic was even worth discussing. The tests were deemed “adequate” as opposed to “too easy” by Boston University Chancellor and State Board of Education Chair John Silber, and Speaker of the House Thomas Finneran described those who failed the tests as “idiots” and characterized teacher preparation programs as education for “dum-dums” (Gorth, 1998).

Manufacturing Acceptability

Barely concealed by the language of quality control, a battle continues to be waged between the Department of Education and the teachers' unions/schools of education. The Department of Education used the poor test results from the first two rounds of test administration to create a number of requirements that would monitor schools of education and improve the "quality" of teacher candidates. The most often cited requirement shuts down schools of education that do not produce a minimum 80% pass rate among teacher candidates taking the state test (Estrin, 1998). Although news stories highlighted these "get-tough" State Board recommendations, pleas by higher education spokespersons for independent analysis of the test were slighted by the media and ignored by state officials. Silber made clear the true intentions of the board of education: "We're not going to have successful reform of education until we destroy the monopoly of teacher colleges in the certification of teachers" (Bradley, 1998, p. 35). The governor at that time, Paul Celluci, agreed: "There's nothing wrong with the test. There's something wrong with the colleges that prepare these people" (Letourneau & O'Shea, 1998, p. A1). However, critics continue to argue that there is something wrong with the tests—that they are flawed in key areas of validity and reliability as testing instruments, that as standardized tests they don't meet recognized psychometric industry guidelines, and that they discriminate against people of color and are culturally biased (Ludlow, 2001). Calls for independent reviews of the test, however, continue to be ignored by the state.

Teacher educators were eventually put on the defensive, which resulted in a number of successes for the state. Because little progress was made in the media to successfully challenge the validity or reliability of the tests, college officials were eventually forced to respond to the criticism of student performance. Although many schools had organized study/review sessions early on and talked about "getting tough" on standards and assessment, the report by the Association of Independent Colleges and Universities of Massachusetts (AICUM) offered a mixed message that may have legitimized the many months of criticism of education schools. The report first forcefully concluded that "no evidence has been provided to the public . . . that the Massachusetts Educator Certification Tests are valid . . . or that these tests are reliable" (Melnick & Pullin, 1999, p. 1). However, the AICUM report also recommended that colleges and universities in the state "reassess their commitment to teacher education and to the activities of both education programs and arts and sciences programs" (p. 36) so that teachers will be prepared "with a deeper understanding of the disciplinary knowledge that undergirds school subject matter and the corollary analytical and productive communications skills that come from something like substantive interaction with disciplinary knowledge" (p. 37). The media then used such concessions to report that "colleges admit failure" in preparing students for the test (Zernicke, 1999). AICUM's goal was to criticize the test while also attempting to describe and endorse many of the efforts to improve teacher preparation in the state, including institutional change. But politically, an acknowledgment that institutions were in any way culpable legitimized the single-minded criticism coming from state government. The

State Board of Education, supported by the governor, much of the state legislature, and the media, had seized the momentum needed to legitimize the test and solidify criticism of higher education in general and education schools in particular.

Given this sort of playing field assessment leaders face, how might we respond in the future? Scholars had worked in the 1990s to build alternative teacher assessment models that challenged minimalist assumptions about teacher proficiency (Delandshere & Petrosky, 1998). The struggle was to create a model that assessed teachers teaching, but the complex and nuanced nature of this movement may be at odds with authorities whose goal is neither complex and meaningful evaluation nor evidence of teaching ability. It is clear from the Massachusetts case that the most fundamental assumptions concerning how and why teachers should be tested need to be addressed. If this doesn't happen, the most deeply flawed test can go through an undeserved process of legitimation. In some ways, the NES writing test had manufactured acceptability from various interests from the beginning. By building redundancy into a system designed to repeatedly measure "correctness" in different ways, the tacit approval of multiple interests could eventually be realized. NES clearly sought to create a literacy test that demanded both the ability to identify errors as well as avoid them through demonstrations of writing. Those who would argue that students must be required to "demonstrate" writing proficiency under "controlled" conditions and be evaluated directly by outside readers were appeased. Those who would complain that students don't know their comma rules and should be assessed "indirectly" through machine-scored tests that require complex sentence manipulations were also content (along with the testing company because this form is so easily evaluated). The test's first priority was thus to satisfy a broad range of the public concerned with correctness and the preservation of a current-traditional rhetorical model. NES also had to create a rigorous, demanding test that pleased its customer, the state board of education, but that could also be implemented with the eventual approval, or at least acquiescence, of the state's teachers and teacher education professionals, since two sections of the writing tests had to be read and evaluated by a cadre of readers from the teaching profession. Once a professional educational organization suggested that teacher education professionals were in some way responsible for low test scores, teachers in higher education may have been more likely to convince themselves that the test itself was acceptable and participate in scoring sessions or test preparation workshops or both. What alternative process could have led to a more progressive model of writing assessment that reflects the work of composition as a discipline over the last 30 years? In Massachusetts, no such process was likely possible given the political players who would have prevented any such model from being created.

Political Associations

The associations of John Silber and other State Board of Education members with conservative educational organizations should also be raised as a major influence in the Massachusetts literacy equation. Four out of nine board members were directly affiliated with the Pioneer Institute, a public policy think tank that advances conservative educational doctrine, especially the privatization of

public education. This creates at least the potential for conflict of interest between those who, in an official capacity, seek to improve public education, while simultaneously working to dismantle it through their unofficial involvement in a private organization (Dunphy, 1999). The Pioneer Institute agenda includes the promotion of charter schools and widespread testing aimed to prove charter schools are effective, even though charter schools in the state have been shown to serve a more affluent student population than local public schools. Narrow ideological orientations and lack of diverse perspectives led Frank Haydu, the commissioner of education who resigned in 1998, to characterize the state board as “dysfunctional and adversarial” with seven of nine members coming from elite higher education backgrounds and no understanding of the different stakeholders in the system (Bradley, 1998).

A Look to the Future and A Call To Action

The objections over the serious irregularities in the creation and initial administration of the Massachusetts teacher test have thus far not effected any change in the administration of the MTEL. The ultimate acceptance, or at least tolerance, of these tests by many higher education professionals as a legitimate requirement in Massachusetts was achieved at least in part through the participation of teachers and college administrators in a media-managed dialogue focused on improving test scores, which tended to suppress dialogue concerning the validity of the test itself. By limiting discussion to the test *scores*, politicians and the media successfully used the familiar literacy-in-crisis argument as an excuse to increase their control over higher education in the state, especially over schools of education, and to successfully eclipse concerns teachers of English and administrators might have for how their students become literate or *what counts* as literacy. Politics counted more than theory and research (Williamson & Huot, 1993).

These examples of high-profile, high-stakes maneuvers by politicians who knew little or nothing about testing literacy, but who played a role in defining who was “literate” and who was not, should dispel any notion that literacy is a fixed, neutral, or objective set of skills, *even though that notion is the basis for the entire argument for the literacy tests*. This chain of events illustrates the highly visible and charged atmosphere of literacy negotiations we are likely to see more of in the future, should we become involved. Those of us involved in assessing writing need to consider how we might best exercise our influence over what counts as valid writing, and do what we can to make our collective voice heard. The Massachusetts test reveals the importance of professional education organizations speaking out against politicking and profiteering as well as asserting their political will. In Massachusetts, organizations like AICUM should have been involved from the beginning in helping to create better forms of assessment with the help of reading and writing experts. But they were given no voice. In this case, state government was accountable to no one. The federal government has historically taken no interest in regulating testing except to allow private companies to do what they will, the sole exception the legislation proposed by the late Paul Wellstone of Minnesota to limit the use of high-stakes tests—an important symbolic action that had no chance

of being enacted. In many states, basic organizing at the grassroots level will be essential if we are to make our voices heard.

What can professional organizations do? NCTE has gone on record as opposing “single issue measurement for the initial credentializing or licensing of teachers and the continuing appointment of teachers,” and has proposed resolutions that support those who oppose high-stakes testing done by private companies and state governments (Board of Directors, 1999). But will resolutions alone change the way reading and writing is assessed in the future? The lessons from Massachusetts underscore the importance of teachers in higher education joining forces with K-12 professionals in order to form political coalitions with existing advocacy groups such as students’ rights groups, teachers’ unions, schools of education, minority rights, and especially those of us who care about who defines literacy, how, and for what purpose(s). The use of high-stakes tests in K-12 schools has resulted in the successful mobilization of teachers, parents, and students from many backgrounds who recognize the tests as a threat to quality public education. We in higher education should recognize the testing of teachers by state governments as potentially dangerous and divisive, and join with our K-12 colleagues in order to influence other organizations and the public. We waste valuable time debating models of assessment among ourselves if in the end those in power don’t listen to us and ignore or undermine any model we propose. Talking to ourselves without acting politically on that talk is pointless if our voices are not heard by those who ultimately make the decisions.

References

- Appleton, J. (1998, July 2). Panel gives no breaks on testing. *Springfield Union News*, p. A1+. Associated Press. (1998, October 18). Colleges unhappy with meeting on tests. *Springfield Union News*, p. A8.
- Associated Press. (1999, July 15). Company defends teacher test. *Springfield Union News*, p. A8.
- Blair, J. (2001, April 18). High noon. *Education Week*. Retrieved May 30, 2001 from <http://www.edweek.org/ew/ewstory>.
- Board of Directors, National Council of Teachers of English. (1999). NCTE resolutions: On high stakes testing. *College English*, 62(3), 415.
- Bradley, A. (1998, December 9). Test questions. *Education Week*, pp. 30-35.
- Bradley, A. (2000, March 22). Washington state to start testing its teachers. *Education Week*. Retrieved May 30, 2001 from <http://www.edweek.org/ew/ewstory>.
- Brinkley, E.H., & R.A. Sudol. (1999). Critical literacy and large-scale writing assessment: The Michigan high school proficiency test. In R. A. Sudol & A. Horning (Eds.), *The literacy connection* (pp. 43-67). Cresskill, NJ: Hampton Press.
- Camp, R. (1996). New views of measurement and new models for writing assessment. In E. White, W.D. Lutz, and S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 135-147). New York: Modern Language Association
- Delandshere, G., & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher*, 27(2), 14-24.
- Delandshere, G., & Petrosky, A. (1999). It is also possible to quantify colors but that is not the point: A rejoinder to Brookhart. *Educational Researcher*, 28(3), 28-30.
- Driscoll, D. (2002, February). *Memorandum: Report on the technical advisory committee for the Massachusetts tests of educational licensure (MTEL)*. Commonwealth of Massachusetts Department of Education.

- Dunphy, P. (1999, June 13). Pioneer thought marks state education trends. *Boston Globe*, final ed., p. E2.
- Educational Testing Service. (1999-2000). Registration bulletin, Praxis Series. Princeton, NJ: Author.
- Estrin, R. (1998, October 13). Teacher education reforms proposed. *Springfield Union News*, p. A1+.
- Fox, T. (1999). *Defending access: A critique of standards in higher education*. Portsmouth, NH: Boynton/Cook/Heinemann.
- Gee, J. (1990). *Social linguistics and literacies: Ideology in discourses*. Brighton, England: Cross-Cultural Approaches to Literacy.
- Gorth, W. P. (1998, December 13). Tests for teachers are right for Massachusetts. *Boston Globe*, third ed. Learning Sec., p. M9.
- Gough, P. B. (1999). For the record. *Phi Delta Kappan*, 80(10), 8 pars. Retrieved October 19, 1999 from http://web2.infotrac.galegroup.com/itw/...!xrn_2_0A55061285?sw_aep+mlln_w_sprcol.html.
- Haney, W., Fowler, C., & Wheelock, A. (1999, February). Less truth than error? An independent study of the Massachusetts Teacher Tests. *Ad Hoc Committee to Test Teacher Test, c/o Center for the Study of Testing, Evaluation, and Education Policy*. Champion Hall, Boston College, Chestnut Hill, MA.
- Hillocks, Jr., G. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Hoff, D.J. (2000, April 5). Appeals court to review California teacher test for bias. *Education Week*. Retrieved on May 30, 2001 from <http://www.edweek.org/ew/ewstory>.
- Holdstein, D.H. (1996). Gender, feminism, and institution-wide assessment programs. In E. White, W. D. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, and practices* (pp. 204-225). New York: Modern Language Association.
- Hood, S., & Parker, L. (1991). Minorities, teacher testing, and recent U.S. supreme court holdings: A regressive step. *Teachers College Record*, 92(4), 603-618.
- Huot, B., & Williamson, M. (1997). Rethinking portfolios for evaluating writing: Issues of assessment and power. In K. Yancey & B. Huot (Eds.), *Assessing writing across the curriculum: Diverse approaches and practices*. Greenwich, CT: Ablex.
- Huot, B., & Williamson, M. (Eds.). (1993). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Cresskill, NJ: Hampton Press.
- Huot, B. (1996, December). Toward a new theory of writing assessment. *College Composition and Communication*, 47(4), 549-566.
- Huot, B., & Schendel, E. (2001). Reflecting on assessment: Validity inquiry as ethical inquiry. *Journal of Teaching Writing*, 17, 37-55.
- Lehman, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus & Giroux.
- Letourneau, N., & O'Shea, M. (1998, August 7). Celluci: System failed students. *Springfield Union News*, p. A1+.
- Ludlow, L. (2001). *Teacher test accountability: From Alabama to Massachusetts*. Educational Policy Analysis Archives, 9.6, Retrieved January 13, 2003 from <http://epaa.asu.edu/epaa/v9n6>.
- Luna, C., Solsken, J., & Kutz, E. (2000). Defining literacy: Lessons from high stakes teacher testing. *Journal of Teacher Education*, 51(4), 276-288.
- Massachusetts Department of Education. (1998). *Massachusetts teacher tests: Test information booklet*. Malden, MA.
- Melnick, S., & Pullin, D. (1999). Teacher education and testing in Massachusetts. Commissioned by The Association of Independent Colleges and Universities of Massachusetts (AICUM). (173 pars). Retrieved March 9, 2000 from the World Wide Web <http://www.masscolleges.org/pf.htm>.
- Melnick, S., & Pullin, D. (2000). Can you take dictation? Prescribing teacher quality through testing. *Journal of Teacher Education*, 51(4), 262-275.

- Moss, P. (1994). Validity in high stakes writing assessment: problems and possibilities. *Assessing Writing*, 1(1): 109-128.
- Pence, L. (1998, July). Saying so will not make it so. *English in Australia*, pp. 56-69.
- Pierce, C. (2003, March 2). Testing times. *Boston Globe Magazine*, pp. 13-22..
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.
- Scanlon, B. (1998, April 6). Teacher tests too easy, official says 95-98 percent of takers pass on first try; board will be asked to toughen standards. *Denver Rocky Mountain News* (CO) local sec.21A. Retrieved July 21, 1999 from <http://infoweb5.newsbank.com/bin/gate.exe?f=doc&state=gq0cmr.3.124>.
- Schaeffer, R. A. (1998, July 15). Testimony of Robert A. Schaeffer, National Center For Fair and Open Testing (Fairtest) Before the Joint Committee on Education. (12 pars.) Retrieved November 8, 1999 from <http://www.fairtest.org/empl/ttest.htm>.
- Schultze, R. (1992, January 27). Teacher testing under fire: State considers dropping professional skills exam. *The Phoenix Gazette* (AZ), p. A1. Retrieved July 21, 1999 from <http://infoweb5.newsbank.com/bin/gate.exe?f=doc&state=gq0cmr.3.84>
- Shaughnessy, M.P. (1977). *Errors and expectations: A guide for the teacher of basic writing*. New York: Oxford.
- Street, B. (1984). *Literacy in theory and practice*. Cambridge, England: Cambridge University Press.
- Street, B. (1995). *Social literacies: Critical approaches to literacy in development, ethnography, and education*. London, England: Longman Books.
- Street, B. (1998). New literacies in theory and practice: What are the implications for language in education? *Linguistics and Education*, 10(1): 1-24.
- Tantraphol, R. (2000, January 27). Driscoll open to test options. *Springfield Union News*, p. A1+.
- Weiss, K.R. (2000, January 16). Rich youth found to get SAT breaks. *Boston Globe*, nation sec., p. B1.
- White, E. (1996). Power and agenda setting in writing assessment. In E. White, W.D. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, and practices*. New York: Modern Language Association.
- White, E. (1998). *Teaching and assessing writing: Recent advances in understanding, evaluating, and improving student performance*. Portland, ME: Calendar Islands Publishers.
- Zernike, K. (1999, April 7). Colleges admit role in failure of teachers. *Boston Globe*, third ed. metro sec., p. A1.