



PROJECT MUSE®

---

## High-Stakes Testing and Students: Stopping or Perpetuating a Cycle of Failure?

Catherine Horn

Theory Into Practice, Volume 42, Number 1, Winter 2003, pp. 30-41  
(Article)

Published by Ohio State University College of Education



➔ For additional information about this article

<https://muse.jhu.edu/article/41414>

# High-Stakes Testing and Students: Stopping or Perpetuating a Cycle of Failure?

*As state-mandated standardized testing becomes an increasingly popular tool by which to make student-level high-stakes decisions such as promotion or graduation from high school, it is critical to look at such applications and their effects on students. Findings in this article suggest that non-White, non-Asian students, as well as students with special needs and English Language Learners, are among the groups most deeply affected by high-stakes testing. Test scores give us important information, but they do not give us all the information necessary to make critical decisions. Given their limited nature and the potentially adverse impacts they can have, using state-mandated large-scale testing for student-level high-stakes purposes is inadvisable.*

USE OF MANDATED LARGE-SCALE TESTING to evaluate programs (e.g., Title 1) has been part of the public educational landscape in the United States for more than 30 years (Heubert & Hauser, 1999). It was the minimum competency era of the 1970s and early 1980s, however, that ushered in the widespread implementation of such tests for student-level evaluations. During this time, the number of such state-level testing programs rose

from 1 in 1972 to 34 by 1985 (Haney, Madaus, & Lyons, 1993). For those states with graduation sanctions attached to their tests, students who were not able to demonstrate minimal competency in the basic skills of reading and arithmetic were denied high school diplomas.

The release of *A Nation at Risk* (National Commission, 1983) reinforced the need for student accountability and elevated the level of demonstrated proficiency. According to the report, the United States could no longer rely on minimal reading and math competency to maintain its competitive edge. Instead, students needed to be held to “rigorous and measurable” standards in order to ensure the country’s success in the information age (National Commission, 1983). These standards would raise the level of expected learning and, in essence, define a new set of minimum competencies. Within 3 years, 35 states had begun comprehensive educational reform, marking the beginning of an almost 2-decade journey to create and hold students accountable for mastery of a new set of world class standards (Kornhaber & Orfield, 2001). Currently, a majority of states use or have plans in place to use state-mandated tests as the sole or significant criteria for promotion and/or graduation from public elementary and secondary schools (“Quality Counts,” 2002).

In order to look at the impacts these state-mandated high-stakes tests have on students, particularly

---

*Catherine Horn is a research associate for The Civil Rights Project at Harvard University.*

those traditionally underserved by the public education system, this article will explore Massachusetts and North Carolina as examples. These states were selected among the 18 currently using high-stakes testing for graduation and/or promotion for several reasons. First, each is in the process of fully implementing standards-based high-stakes testing for graduation and/or promotion. In North Carolina, third, fifth, and eighth grade students are required to pass a state-mandated test for promotion, and a high school exit test will be phased in over the next several years. In Massachusetts, the graduating class of 2003 will be the first group that must pass the state's standards-based exams at the 10th grade in order to graduate from high school. Although many states already fully implement promotion or graduation testing and a majority are moving toward standards-based assessments, a fair number have not yet fully aligned their tests with the state content standards in the way that Massachusetts and North Carolina have. Thus, they are two of the better state-level examples of the aforementioned standards-based reform movement in that they not only hold students accountable, but do so in relation to a prespecified set of rigorous content standards. Additionally, Massachusetts and North Carolina have readily available disaggregated test score data as well as information about the content standards measured on the state exams. Most states provide information on passing rates but often do not present them broken out by race.<sup>1</sup> Likewise, many states make practice questions publicly available, but the specific learning standards measured by the test are not clearly identified. Massachusetts and North Carolina, then, offer particularly good opportunities to disentangle the effects of state-mandated high-stakes tests on minority, ESL, low socioeconomic status, and other students, as well as to discern the kinds of content standards that are being measured.

This article begins by looking at what the research reveals about high-stakes testing and its relationship to student outcomes. It next presents data from Massachusetts and North Carolina on state trends related to high-stakes testing and students. Finally, the author offers some suggestions on the appropriate uses of testing for educational decision making.

## **What Research Says About the Impact of High-Stakes Testing on Students**

As state-mandated standardized testing becomes an increasingly popular tool by which to make student-level high-stakes decisions such as promotion or graduation from high school, it is critical to look at what the literature tells us about such applications and their effects. Although space does not allow for a detailed review, this section highlights some of the major findings across studies.

### **Disparities in performance**

Much work has been done to document and analyze the performance gaps between Whites and Asians relative to Hispanics and African Americans on both tests in general and high-stakes tests in particular. Generally, studies have found that although differences in test scores have narrowed over time, substantial disparities still exist. For example, Hedges and Nowell (1998) found that African Americans have been greatly underrepresented among the highest test scorers on standardized tests, and that underrepresentation has not diminished over time. Similarly, Madaus and Clarke (2001) document that, based on 1996 National Assessment of Educational Progress (NAEP) scores, the average proficiency for White 13-year-olds was about the same level achieved by 17-year-old African Americans, and that Hispanics also continue to underperform relative to their White counterparts.

Analyses of test score differences between regular education students and students classified in special groups, such as English Language Learners (ELLs) or students with disabilities, also show that without appropriate accommodations (and sometimes even with them), the latter two typically underperform (DeStefano, 1998; LaCelle-Peterson, 1998).<sup>2</sup> For example, the work of McNeil and Valenzuela (2001) found that children in Texas with limited English proficiency were being "especially handicapped in their ability to exhibit their knowledge by the [Texas Assessment of Academic Skills] TAAS exit test" (p. 147). Students with disabilities in New York, regardless of the type of accommodations received, still greatly underperformed on the Regents exams relative to their nondisabled counterparts (Koretz & Hamilton, 2001). Such documented disparities must be carefully considered when weighing these tests' impacts.

### Dropout rates

Determining the impact high-stakes testing may have on dropout rates is complicated. The confounding influences of factors ranging from the end of social promotion, to immigrant status, to changes in graduation requirements make it difficult to pinpoint a single influence as the root cause of a student's decision to leave school before graduating. That said, a growing body of research is attempting to more clearly disentangle the impact of high-stakes exit testing on dropping out (Heubert & Hauser, 1999). There is some empirical work to argue that no relationship exists between high-stakes testing and dropping out (Bishop & Mane, 2001). There is a larger body of research, however, that suggests such exit tests are related to an increase in the numbers of students dropping out, particularly for students already at risk (Catterall, 1989; Kreitzer, Madaus, & Haney, 1989; Madaus & Clarke, 2001). In one of the most recent large-scale studies on this issue, Haney (2000) studied the impact of the TAAS on school completion in Texas and found evidence to suggest that the exit exam was associated with an increase in dropout rates, especially among African Americans and Hispanics. High-stakes testing, then, may increase the numbers of students leaving high school without a diploma—a minimum certification necessary in today's labor market.

### Retention rates

Districts and states are beginning to increasingly rely on mandated high-stakes tests to make promotion decisions. Chicago Public Schools is perhaps the most well known of such a district-level implementation, where policy makers mandated minimum performance on the Iowa Test of Basic Skills (ITBS) in order to be promoted. In its first year, 15%, 13%, and 8% of the students at grades 3, 6, and 8, respectively were retained based on ITBS test scores, even after mandatory summer school (Heubert & Hauser, 1999). While Chicago represents only district-level data, findings such as these highlight key issues that states may face on an aggregated level. Further, the statewide use of high-stakes testing for promotion decisions has to be considered in the broader context of current retention trends in which a large share of American

school children are already retained (Heubert & Hauser, 1999). Again, minorities, ELLs and students with disabilities are likely to be the most vulnerable to such policies (Shepard, 1991). In considering whether retention is detrimental in and of itself, existing research once again provides some insight. Data indicate that repeating a grade generally does not improve achievement, and it often increases the dropout rate (Heubert & Hauser, 1999; Shepard & Smith, 1989).

### Student learning

The research shows that the negative impacts of high-stakes testing on students are potentially severe. But we have not addressed a possibly redeeming factor: whether the exams serve their intended purpose of improving student learning. As the argument goes, high-stakes tests "focus student attention on the knowledge and skills that are deemed most important to learn" (Linn & Herman, 1997, pp. 2, 5). This important knowledge and skill set, however, often becomes myopically defined as a narrow, test-defined set of skills (Madaus & Clarke, 2001). Students focus on mastering only those competencies measured on the exam to the exclusion of others that may be educationally important but untested (e.g., collaboration, research project design). Teachers foster those efforts by teaching to the content and tradition of the test (Madaus, 1988). Test scores go up and more students pass the exam.

As empirical evidence suggests, however, increased high-stakes test scores do not equate to increased learning (Cannell, 1989; Koretz, Mitchell, & Stetcher, 1996). For example, work by researchers at RAND found that, while TAAS scores in Texas indicated large increases in academic achievement across all ethnoracial groups, NAEP scores during the same time period suggested otherwise. While Texas students improved significantly more on a fourth-grade NAEP math test than did their counterparts nationally, the size of this gain was smaller than their gains on TAAS. Further, such gains were not present on the eighth-grade math test. In particular, where TAAS scores suggested a rapid narrowing of the achievement gap between Whites and students of color, NAEP trends showed a larger and increasing gap (Klein,

Hamilton, McCaffrey, & Stetcher, 2000). These findings suggest that high-stakes tests are not necessarily leading to increased learning. Similarly, Amrein and Berliner (2002) gathered comprehensive evidence from 18 states using high-stakes testing to suggest that in all but one analysis, student learning was indeterminate, remained at the same level as before the policy was implemented, or actually went down after the testing policy was instituted. These high-stakes tests, then, may be increasing risks with no increased benefits to student learning.

Given this general synopsis of the ways in which high-stakes testing may impact students, the article now turns to Massachusetts and North Carolina as specific examples of how state-mandated high-stakes exams are affecting students.

### Massachusetts

In 1993 the Massachusetts Education Reform Act was passed to ensure that all students were learning at high levels. Groups of educators began the task of creating frameworks “of high quality, results driven, and focused on world class standards” (Massachusetts Curriculum Frameworks, n.d., para. 3). To assess whether students were meeting those expectations, the Massachusetts Comprehensive Assessment System (MCAS) was created and administered for the first time in 1998. Passing scores on the exams would be an indication that test takers could “synthesize, organize, and apply knowledge to complex problems and real-life situations” (“Background on the MCAS,” n.d., para. 2 and following).

In its current iteration, the MCAS is administered at grades 3, 4, 5, 6, 7, 8, and 10. Across these grades, tests in reading, English Language Arts (ELA), mathematics, science and technology/engineering, and history and social science are administered. They include multiple-choice, short answer, and open-response items. Students receive a scaled score (ranging from 200 to 280) and a corresponding proficiency level: *Warning*, *Needs Improvement*, *Proficient*, or *Advanced*. For students in the graduating class of 2003, a *Needs Improvement* or better on both the ELA and mathematics exams is necessary to graduate from high school.

### MCAS results for the 10th grade

Tables 1 and 2 present the percentages scoring at the *Needs Improvement* level or higher on the 10th grade ELA and mathematics tests for each of the years the exams have been administered. Overall, more students pass the ELA MCAS than do the mathematics MCAS. Minority students, in particular African Americans and Hispanics, however, greatly underperform relative to their White and Asian counterparts on both tests. For example, only 52% and 42% of Hispanic ELA and mathematics test takers, respectively, scored at the *Needs Improvement* level or higher on the 2001 administration. Comparatively, 88% of White students reached the same level on the ELA test; 82% did so on the math MCAS. Regular education students outperform students with disabilities and limited English proficient students every year on both tests.

For the first three years of administration, the percentages passing the ELA test stayed fairly stagnant or declined across racial/ethnic and special population categories. The mathematics test scores reflected similar patterns in the first two years, but the percentage passing actually rose modestly (3 to 7 percentage points) across all categories in the Spring 2000 administration. In 2001, the first year test takers needed a *Needs Improvement* score on both tests to graduate, however, all racial/ethnic and student status groups saw marked increases in the percentages passing each of the exams. Although there is no clear and definitive explanation for these substantial jumps, several possibilities exist. Students might have taken the tests more seriously because of the upcoming diploma sanction. Additionally, schools may have worked to improved student performance on the tests by offering more focused in-class and after-school preparation for the exams.

Alternatively, some have argued that the gains may not be due to increased student performance but to technical changes in the way raw scores were converted to scaled scores (Hayward, 2001). Skeptics of the increase have also pointed to the fact that more than 4,000 fewer students are present in the current class of 2003 compared with the original class of 2003, following patterns seen in the research on high-stakes testing, retention, and dropping out in Texas (Haney, 2000). A state report

Table 1  
Percent Meeting the Needs Improvement Standard or Higher on the 10th Grade ELA MCAS

|              | African Americans | Asians | Hispanics | Mixed Race | Native Americans | Whites | Regular Education Students | Students with Disabilities | Limited English Proficient Students |
|--------------|-------------------|--------|-----------|------------|------------------|--------|----------------------------|----------------------------|-------------------------------------|
| Spring, 1998 | 51                | 74     | 42        | 70         | 43               | 81     | 78                         | 33                         | 36                                  |
| Spring, 1999 | 43                | 70     | 37        | 65         | 41               | 75     | 76                         | 27                         | 31                                  |
| Spring, 2000 | 40                | 66     | 34        | 60         | 35               | 73     | 66                         | 25                         | 25                                  |
| Spring, 2001 | 60                | 80     | 52        | 79         | 70               | 88     | 82                         | 47                         | 38                                  |

Sources: Massachusetts Department of Education (1998, 1999, 2001)

Table 2  
Percent Meeting the Needs Improvement Standard or Higher on the 10th Grade Mathematics MCAS

|              | African Americans | Asians | Hispanics | Mixed Race | Native Americans | Whites | Regular Education Students | Students with Disabilities | Limited English Proficient Students |
|--------------|-------------------|--------|-----------|------------|------------------|--------|----------------------------|----------------------------|-------------------------------------|
| Spring, 1998 | 20                | 60     | 17        | 43         | 21               | 57     | 54                         | 12                         | 18                                  |
| Spring, 1999 | 20                | 60     | 15        | 45         | 23               | 55     | 53                         | 13                         | 4                                   |
| Spring, 2000 | 23                | 67     | 21        | 48         | 34               | 62     | 61                         | 17                         | 23                                  |
| Spring, 2001 | 48                | 85     | 42        | 70         | 55               | 82     | 82                         | 38                         | 43                                  |

Sources: Massachusetts Department of Education (1998, 1999, 2001)



suggests that such drops are normal and are typically the result of students dropping out (unrelated to the test), being retained in grade (unrelated to the test), or moving out of the state (Perlman, 2002). In order to more fully understand whether the MCAS graduation exams, other factors, or some combination of the two may be resulting in more dropouts and retentions in grade, the state must continue to closely document and make publicly available the disaggregated data on who is being impacted and why.

Turning back to the current class of 2003, the first group to fall under the high-stakes testing stipulation, Table 3 shows the numbers of students by racial/ethnic or student status category who have earned a competency determination as of the Fall 2001 retest.<sup>3</sup> According to the state, the current class of 2003 is comprised of 8% African Americans, 5% Asians, 8% Hispanics, and 79% Whites. Of that total, 76% of the students have passed the necessary exams to graduate. Disaggregated by race/ethnicity, however, the total number having earned a competency determination is made up of 5% African Americans, 5% Asians, 4% Hispanics, and 85% Whites. African Americans and Hispanics are underrepresented relative to their presence in the enrolled class; Whites are noticeably overrepresented. Likewise, students with disabilities and limited English proficiency make up 12% and 4%, respectively, of the Fall 2001 enrollment but only 7% and 1% of the total having earned a competency determination.

The disproportionate impact of the 10th grade MCAS tests is presented differently in Table 4. While 82% of all White students in the current class of 2003 have met the testing requirements necessary for graduation as of the Fall 2001 retest, only 41% of Hispanics and 48% of African Americans have met the same goal. Even more striking, 84% of limited English proficient students have not yet passed both tests. Despite the large increases in the percentages passing (as shown in Tables 1 and 2), it seems unlikely that the gains necessary to put minority pass rates on par with Whites and Asians will be met with only one year remaining until the 2003 graduation date.

### North Carolina

Approved by the state in 1999, the North Carolina Board of Education passed sweeping reforms

**Table 3**  
**Percent of Total by Category Enrolled**  
**Compared with Percent of Total by**  
**Category Who Have Earned a Competency**  
**Determination, Fall 2001**

|                               | Fall 2001<br>Enrollment<br>(% of Total<br>Enrollment) | Total Number Who<br>Have Earned CD<br>(% of Total Who<br>Have Earned CD) |
|-------------------------------|---|--|
| All Students                  | 63767   | 48463  |
| Limited English<br>Proficient | 2335<br>(4)   | 374<br>(1)   |
| Students with<br>Disabilities | 7806<br>(12)  | 3512<br>(7)  |
| Regular<br>Education          | 53626<br>(84)   | 44510<br>(92)  |
| African Ameri-<br>can/Black   | 5074<br>(8)   | 2436<br>(5)  |
| Asian                         | 2956<br>(5)   | 2217<br>(5)  |
| Hispanic                      | 5284<br>(8)   | 2166<br>(4)  |
| Native American               | 172<br>( $<1$ )                                       | 126<br>( $<1$ )  |
| White                         | 50281<br>(79)   | 41230<br>(85)  |

Source: Massachusetts Department of Education (2002)

**Table 4**  
**Percent of Total Meeting the Needs**  
**Improvement Standard or Higher on both**  
**the 10th Grade ELA and Mathematics**  
**MCAS, Fall 2001**

|                            | % of Total Who<br>Have Earned CD |
|----------------------------|----------------------------------|
| All Students               | 76                               |
| Limited English Proficient | 16                               |
| Students with Disabilities | 45                               |
| Regular Education          | 83                               |
| African American/Black     | 48                               |
| Asian                      | 75                               |
| Hispanic                   | 41                               |
| Native American            | 73                               |
| White                      | 82                               |

Source: Massachusetts Department of Education (2002)

requiring increased accountability at the student level. Although the state already had structures in place to evaluate schools (e.g., The ABCs of Public Education), these new reforms marked the first time the state implemented high-stakes testing for elementary and middle school students. Students receive a scaled score and a corresponding achievement level ranging from I (student does not have sufficient mastery to be promoted) to IV (student performs beyond what would be expected to be promoted) (North Carolina State Board of Education, 1999).

Students in grades 3 through 8 participate in Gateway Exams, a set of standards-based, end-of-grade multiple-choice tests of reading and mathematics, in order to be promoted from grades 3, 5, and 8. High school students, beginning with the graduating class of 2005, will also have to pass an exit exam to graduate from high school. In order to “ensure that students are working at grade level in reading, writing, and mathematics before being promoted to the next grade,” fifth graders in 2000-2001 were the first required to pass reading and math Gateway Exams for promotion (North Carolina Statewide Student, n.d., para. 1). The exam-based retention policies were applied to third and eighth graders with the 2001-2002 test administrations (data on the number of students retained as a result of test scores are not yet available). It is important to note, however, that principals have overriding power to make promotion and retention decisions (North Carolina State Board of Education, 2000).

**Results for the fifth grade**

Table 5 presents the percentages of fifth graders scoring at or above Level III on both the reading and mathematics tests over a 2-year period. As evidenced in the table, all racial/ethnic groups have seen noticeable increases. The percentage of African

Americans passing both tests, for example, has risen 6 points. Substantial disparities, however, remain among African Americans, Hispanics, Asians, and Whites. While 87% of White and 85% of Asian students passed both Gateway Exams in 2001, only 62% and 67% of African Americans and Hispanics, respectively, met the same standard. These differential performances among racial/ethnic groups are substantive, especially given the fact that Spring 2001 marked the first year the fifth grade exams were used for promotion decisions.

While Table 5 presents the percentages of students passing both the reading and math Gateway Exams on the first try, it does not, according to other state documents, exactly reflect the percentage of students promoted to the sixth grade. In determining how many students are promoted/retained in grade based on end-of-grade test scores, North Carolina offers two retests (the first of which happens only 3 weeks after the first test results are returned) to students not meeting the requirements on the first exam. Additionally, students who do not meet the standards on the second try receive a Personalized Education Plan (PEP), which is intended to provide focused instruction. Further, the state gives schools the option of adding one standard error of measurement to a student’s test scores in determining whether (s)he has met the designated achievement level for passing (Student Accountability Standards, n.d.). Finally, as mentioned earlier, there is a review process in place that allows principals to make a decision to promote a student despite not reaching the standards (North Carolina State Board of Education, 2000).

Taking all of those safeguards into place, then, 92% of the state’s tested fifth graders passed both sections and were promoted in 2001 (North Carolina State Board of Education, 2001). Conversely, less

**Table 5**  
**Percent Scoring At or Above Level III in Both Reading and Mathematics, Grade 5**

|              | African Americans | Asians | Hispanics | Mixed Race | Native Americans | Whites |
|--------------|-------------------|--------|-----------|------------|------------------|--------|
| Spring, 2000 | 56                | 82     | 62        | 77         | 59               | 83     |
| Spring, 2001 | 62                | 85     | 67        | 82         | 65               | 87     |

Source: North Carolina State Board of Education, State Testing Results, 2000-2001



than 3% of the tested fifth graders did not meet the standards and were retained. Five percent of the fifth graders did not meet the standards but were promoted; roughly 1% met the standards but were not promoted for other reasons. Among the various racial/ethnic groups, Whites and Asians had the highest percentage meeting the standard and being promoted (96% and 94%, respectively). Hispanics and African Americans had the lowest percentage meeting the standards and being promoted (85% and 86%, respectively). Four percent of both African American and Hispanic test takers were retained in grade because they did not meet the standards (compared with 1% of all White students) (North Carolina State Board of Education, 2001).

Taken as a whole, the Massachusetts and North Carolina results suggest that non-White, non-Asian students are among the groups most affected by this type of high-stakes testing. The 10th grade results from the MCAS ELA and mathematics exams show that minority, limited English proficient, and disabled students will be deeply impacted by the upcoming diploma sanction. As many as half of African Americans and Hispanics currently in the class of 2003 may not graduate because of test scores. Up to 84% of limited English proficient students also may not receive a diploma. In North Carolina, African American and Hispanic students are being retained in grade because of test scores at almost 4 times the rate of White and Asian students. Although it is too early to determine the extent to which these high-stakes tests are having other negative impacts on students (e.g., increased dropout rates), preliminary descriptive data are troubling.

But the question remains, in spite of these findings. Are such tests valuable because they ultimately ensure an education that will better prepare students for the needs of a changing workforce? The next section addresses this issue directly.

### **21st Century Skills and a Changing Workforce**

In 1999, the U.S. Departments of Commerce, Education, and Labor, along with the National Institute of Literacy and the Small Business Administration, released a report outlining 21st century skills necessary for 21st century jobs. Those competencies include:

- Basic skills – The academic basics of reading, writing, and computation are necessary for jobs of all kinds.
- Technical skills – Workers use a growing array of advanced information, telecommunications, and manufacturing technologies, as employers turn to technology to boost productivity and efficiency, and to deliver services to customers in new ways.
- Organizational skills – New systems of management and organization, as well as employee-customer interactions, require a portfolio of skills in addition to academic and technical skills. These include communication, analytical, problem solving, and interpersonal skills; creative thinking; and the ability to negotiate and influence and to self-manage.

Others have similarly laid out the skills students ought to have to be productive citizens in the 21st century (CEO Forum on Technology and Education, 2001; Murnane & Levy, 1996). As evidenced by the list above and elsewhere, workers in the emerging labor market will be required to have far more than the basic skills of reading, writing, and math. The recurring skills needed for success in the new millennium also include proficiencies in technology, communication, problem solving, and working with others. Moreover, these 21st century skills are demanded in the context of a changing labor force. The U.S. Department of Labor projects that by 2008 Asians, Hispanics, and African Americans will have a 40%, 37%, and 20% increased presence in the labor force, respectively. Whites, by comparison, are estimated to have only a 7% increase (U.S. Dept. of Labor, n.d.).

While promotion tests are seen as important preventive measures, high school exit exams, in particular, are perceived as a last stop gap for underprepared students funneling into the workforce. Given their more direct link to the labor market, then, the article returns to the 10th grade MCAS.

### **ELA and Math curriculum frameworks, the MCAS, and 21st century skills**

By their own claim, the Massachusetts Frameworks are world class and prepare students to enter the workforce in the new millennium. But how do the standards in the Massachusetts Frameworks

really compare to the necessary 21st century skills just outlined? At present, the 10th grade ELA Frameworks include 27 standards broken into 4 strands: Language, Reading and Literature, Composition, and Media (Massachusetts Department of Education, 2001). The Mathematics Frameworks cover 5 strands: Number Sense and Operations; Patterns, Relations, and Algebra; Geometry; Measurement; and Data Analysis, Statistics, and Probability (Massachusetts Department of Education, 2000).<sup>4</sup> The standards in both frameworks, but in particular the ELA Frameworks, ask students to demonstrate both basic and higher-order thinking skills, work in groups, analyze and apply technology, and so on, and are very much in keeping with those outlined as necessary for the 21st century workforce (Achieve, 2001).

Given that the standards are in line with 21st century competencies, one must next ask how well the MCAS tests assess students' acquisition of those skills. Although the MCAS tests ask only questions aligned with the standards and include short answer and open-response questions in addition to multiple-choice items, not all standards and not all modes of demonstration receive equal attention. Looking specifically at the 10th grade ELA MCAS test, only 18 of the 27 standards across three of the strands (the Media content strand is not included on the test) are measured, and the majority of the questions are in multiple-choice format. The standards *not* tested include the following examples:

- Students will pose questions, listen to the ideas of others, and contribute their own information or ideas in group discussions or interviews in order to acquire new knowledge. (Language Strand)
- Students will make oral presentations that demonstrate appropriate consideration of audience, purpose, and the information to be conveyed. (Language Strand)
- Students will organize ideas in writing in a way that makes sense for their purpose. (Composition Strand)
- Students will design and create coherent media productions (audio, video, television, multimedia, Internet, emerging technologies) with a clear controlling idea, adequate detail, and appropriate consideration of audience, purpose and medium. (Media Strand)

Comparing these nontested ELA standards to the 21st century competencies laid out previously, it is interesting to note that almost all of these harder to measure skills are viewed as no less necessary for jobs in the new millennium. For as much as the new labor market will need to have the ability to read and write, these workers will also need to work collectively, utilize technology, and be able to present ideas orally (to name a few). Similarly, while many of the mathematics standards are more easily and readily tested, there still remain less easily measured skills that are important in the emerging labor market but not assessed (e.g., an ability to express mathematical concepts clearly to a variety of audiences).

At best, the high-stakes MCAS tests are ensuring proficiency in only a subset of skills defined as essential for work in the new millennium. At worst, the MCAS assessments may be leading to the underpreparation of students for the 21st century workforce. The American Educational Research Association (AERA) states, "The content of the test and the cognitive processes engaged in taking the test should adequately represent the curriculum. High-stakes tests should not be limited to that portion of the relevant curriculum that is easiest to measure" (2000, para. 10). Looking to the 10th grade MCAS ELA test as an example, however, the test developers seem to have done just what AERA warns against. Only the most readily testable standards are included on the assessment. Given that research suggests content measured on high-stakes tests ultimately defines the curriculum, valuable skills may be lost because they are not tested and therefore not taught. This is even more troubling in the context of the disproportionately high rates of failure among African Americans, Hispanics, limited English proficient students, and students with disabilities. An increasingly diverse workforce may not be ready for what it will be asked to do.

### Using Tests Wisely

Many professional educational organizations have spoken out strongly against the use of a single test score for promotion and/or graduation of students. The American Evaluation Association (AEA) recently released a position paper stating,

“High-stakes testing leads to under-serving or miss-serving all students, especially the most needy and vulnerable, thereby violating the principle of ‘do no harm’” (2002, para. 1). Basing its position on the 1999 *Standards for Educational and Psychological Testing*, the AERA writes, “Decisions that affect individual students’ life chances or educational opportunities should not be made on the basis of test scores alone” (AERA, 2000, para. 6).

Rather than rely on a single measure, other relevant information such as grades and teacher recommendations should be considered in determining promotion or graduation (Heubert & Hauser, 1999). Such a student-level accountability model balances test performance with other indicators of achievement and allows one measure to offset another. Further, given the disparate test score performance among Whites, African Americans, and Hispanics documented earlier, using a compensatory system seems increasingly logical. As Jencks (1998) argues, test score differences between Whites and minorities may be real, “But inability to measure the other predictors of performance, on which Blacks [and Hispanics] seem to be far less disadvantaged, poses a huge social problem” (p. 84). If tests are not assessing certain qualities indicative of future professional success, it seems advisable to decrease (not increase) reliance on them.

### Conclusion

With the introduction of his education reform initiative, President George W. Bush (2001) outlined a referendum on public education that included the following mandate:

Too much precious time has lapsed in this case for us to achieve what we want: every child being able to learn. Testing every child every year is the way to stop the cycle. We must care enough to ask how our children are doing. (G.W. Bush, press conference, January 2001)

As Bush’s vision has come to fruition with the January 2002 signing of the reauthorized Elementary and Secondary Education Act, standardized testing is now federally blessed to remain a linchpin of educational accountability. It remains to be seen, however, whether the use of standardized tests will lead to every child being able to learn. Madaus and Horn (2000) note, “Although

the use of standardized tests was intended to assist in the improvement of public education and in many ways it has, it also created long-term, intractable problems related to misuse or overuse” (p. 49). Test scores give us important information, but they do not give us all the information necessary to make critical decisions. Given their limited nature and the potentially adverse impacts they can have as evidenced in the literature and in Massachusetts and North Carolina, using state-mandated large-scale testing as the single measure for student-level high-stakes purposes is unadvisable.

### Notes

1. One of the mandates to states in the reauthorized Elementary and Secondary Education Act is to make such data available to the public.
2. For a thorough discussion of the technical and validity issues related to testing students with disabilities and English Language Learners, see Heubert and Hauser (1999).
3. It is important to note that only one focused retest form is administered per year. The Department of Education writes, “Students who have not yet earned a competency determination are allowed to participate in the spring MCAS administration, but must answer the same common questions as students taking the standard test. In the future, an additional focused retest opportunity may be offered in the spring or at some other point during the year” (“Frequently Asked Questions,” n.d., final para.). Additionally, the Fall 2001 retest did not include advanced questions, but instead added more questions focused at the level of *Needs Improvement* and was targeted at students who had not yet met the standard on the ELA and/or math MCAS.
4. A revised version of the English Language Arts Frameworks was released in June 2001. The new frameworks represent a refinement of the original released in February 1997 but remain substantively similar. The 2000 Mathematics Frameworks are a more substantive revision of the 1996 original. To see both of the original frameworks, visit the Massachusetts Department of Education website at <http://www.doe.mass.edu/frameworks/archive/>.

### References

- Achieve, Inc. (2001). *Measuring up: A report on education standards and assessments for Massachusetts*. Cambridge, MA: Author.
- American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*.

- Washington, DC: American Educational Research Association.
- American Educational Research Association. (2000, July). *AERA position on high stakes testing*. Retrieved April 1, 2002, from <http://www.aera.net/about/policy/stakes.htm>
- American Evaluation Association. (2002, February). *American Evaluation Association position statement on high stakes testing In preK-12 education*. Retrieved March 30, 2002, from <http://www.eval.org/hst3.htm>
- Amrein, A., & Berliner, D. (2002, March 28). High stakes testing, uncertainty, and student learning. *Educational Policy Analysis Archives* 10(18). Retrieved May 8, 2002, from [http://epaa.asu.edu/epaa/Background on the MCAS tests of May 1998](http://epaa.asu.edu/epaa/Background%20on%20the%20MCAS%20tests%20of%20May%201998). (n.d.). Malden, MA: Massachusetts Department of Education. Retrieved April 30, 2002, from <http://www.doe.mass.edu/mcas/1998/bg/default.html>
- Bishop, J., & Mane, F. (2001). The impacts of minimum competency exam graduation requirements on college attendance and early labor market success of disadvantaged students. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high stakes testing in public education* (pp. 51-83). New York: Century Foundation.
- Cannell, J. (1989). *The "Lake Wobegon" report: How public educators cheat on standardized achievement tests*. Albuquerque, NM: Friends for Education.
- Catterall, J.S. (1989). Standards and school dropouts: A national study of tests required for graduation. *American Journal of Education*, 98(1), 1-34.
- CEO Forum on Education and Technology. (2001, June). *Key building blocks for student achievement in the 21st century: Assessment, alignment, accountability, access, and analysis*. Washington, DC: Author. Retrieved April 18, 2002, from <http://www.21stcenturyliteracy.org/workplace/>
- DeStefano, L. (1998). High stakes testing and students with handicaps: An analysis of issues and policies. In R. Stake (Ed.), *Advances in program evaluation*, 1 (pp. 267-288). Greenwich, CT: JAI Press.
- Frequently asked questions about the Spring 2002 MCAS administration*. (n.d.). Malden, MA: Massachusetts Department of Education. Retrieved May 15, 2002, from <http://www.doe.mass.edu/mcas/2002/admin/faq.html#V>
- Haney, W. (2000, August 19). The myth of the Texas miracle. *Educational Policy Analysis Archives*, 8(41). Retrieved May 13, 2002, from <http://epaa.asu.edu/epaa/v8n41/>
- Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kluwer.
- Hayward, E. (2001, October 16). Dramatic improvement in MCAS scores. *The Boston Herald*. Retrieved April 15, 2002, from [http://www.bostonherald.com/news/local\\_regional/mcas10162001.htm](http://www.bostonherald.com/news/local_regional/mcas10162001.htm)
- Hedges, L., & Nowell, A. (1998). Black-white test score convergence since 1965. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 149-181). Washington, DC: Brookings Institution Press.
- Heubert, J., & Hauser, R. (Eds.). (1999). *High stakes: Testing for tracking, promoting, and graduation*. Washington, DC: National Academy Press.
- Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 55-85). Washington, DC: Brookings Institution Press.
- Klein, S., Hamilton, L., McCaffrey, D., & Stetcher, B. (2000). *What do test scores in Texas tell us?* [Issue paper]. Santa Monica, CA: RAND.
- Koretz, D., & Hamilton, L. (2001). *The performance of students with disabilities on the New York Regents comprehensive examination of English* (CSE technical report 540). University of California, Los Angeles: Center for the Study of Evaluation. Retrieved on April 15, 2002, from <http://www.cse.ucla.edu/CRESST/pages/reports.htm>
- Koretz, D., Mitchell, B., & Stetcher, B. (1996). *The perceived effects of the Kentucky instructional results information system* (MR- 792-PCT/FF). Santa Monica, CA: RAND.
- Kornhaber, M., & Orfield, G. (2001). High-stakes testing politics. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high stakes testing in public education* (pp. 1-18). New York: Century Foundation.
- Kreitzer, A., Madaus, G., & Haney, W. (1989). Competency testing and dropouts. In L. Weis, E. Farrar, & H. Petrie (Eds.), *Dropouts from schools: Issues, dilemmas, and solutions* (pp. 129-152). Albany: State University of New York Press.
- LaCelle-Peterson, M. (1998). Choosing not to know: How assessment policies and practices obscure the education of language minority students. In A. Filer (Ed.), *Assessment: Social practice and social product* (pp. 27-42). London: Routledge Falmer.
- Linn, R., & Herman, J. (1997, February). *Standards-led assessment: Technical and policy issues in measuring school and student progress* (CSE technical report 426). University of California, Los Angeles: Center for the Study of Evaluation.
- Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum, 87th yearbook of the national society for the study of education* (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G., & Clarke, M. (2001). The impact of high-stakes testing on minority students. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high stakes testing in public education* (pp. 85-106). New York: Century Foundation.
- Madaus, G., & Horn, C. (2000). Testing technology: The need for oversight. In A. Filer (Ed.), *Assessment:*



- Social practice and social product* (pp. 47-66). London: Routledge Farmer.
- Massachusetts curriculum frameworks. (n.d.). Malden, MA: Massachusetts Department of Education. Retrieved April 15, 2002, from <http://www.doe.mass.edu/frameworks>
- Massachusetts Department of Education. (1998, November). *Report of 1998 statewide results: The Massachusetts Comprehensive Assessment System (MCAS)*. Malden, MA: Author. Retrieved April 30, 2002, from <http://www.doe.mass.edu/mcas/results.html>
- Massachusetts Department of Education. (1999, November). *The Massachusetts Comprehensive Assessment System (MCAS): Report of 1999 statewide results*. Malden, MA: Author. Retrieved April 30, 2002, from <http://www.doe.mass.edu/mcas/1999/results/99mcas/110899com.html>
- Massachusetts Department of Education. (2000, November). *Massachusetts mathematics curriculum frameworks*. Malden, MA: Author. Retrieved April 30, 2002, from <http://www.doe.mass.edu/frameworks/current.html>
- Massachusetts Department of Education. (2001, June). *Massachusetts English language arts curriculum frameworks*. Malden, MA: Author. Retrieved April 30, 2002, from <http://www.doe.mass.edu/frameworks/current.html>
- Massachusetts Department of Education. (2001, October). *Spring 2001 MCAS tests: State results by race/ethnicity and student status*. Malden, MA: Author. Retrieved May 5, 2002, from <http://www.doe.mass.edu/mcas/results.html>
- Massachusetts Department of Education. (2002, April). *Progress report on the class of 2003: Percentage of students who have earned competency determination statewide and by district*. Malden, MA: Author. Retrieved April 30, 2002, from <http://www.doe.mass.edu/mcas/results.html>
- McNeil, L., & Valenzuela, A. (2001). The harmful impacts of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In M. Kornhaber & G. Orfield (Eds.), *Raising standards or raising barriers: Inequality and high stakes testing in public education* (pp. 127-150). New York: Century Foundation.
- Murnane, R., & Levy, F. (1996). *Teaching the new basic skills: Principles for educating children to thrive in a changing economy*. New York: Free Press.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington DC: U.S. Department of Education. Retrieved April 20, 2002, from <http://www.ed.gov/pubs/NatAtRisk/risk.html>
- North Carolina State Board of Education. (n.d.). *Personalized Education Plans Training Manual*. Raleigh: Author. Retrieved August 23, 2002, from [http://www.ncpublicschools.org/student\\_promotion/pepmanual/](http://www.ncpublicschools.org/student_promotion/pepmanual/)
- North Carolina State Board of Education. (n.d.). *Student accountability standards implementation guide*. Raleigh, NC: Author. Retrieved July 25, 2002, from [http://www.dpi.state.nc.us/student\\_promotion/SAS\\_guide/contents.html](http://www.dpi.state.nc.us/student_promotion/SAS_guide/contents.html)
- North Carolina State Board of Education. (1999a). *North Carolina Standard Course of Study*. Raleigh, NC: Author. Retrieved July 30, 2002, from <http://www.ncpublicschools.org/curriculum/>
- North Carolina State Board of Education. (1999b). *State of the state: Education performance in North Carolina in 1999*. Raleigh, NC: Author. Retrieved August 23, 2002, from <http://www.ncpublicschools.org/Accountability/reporting/sosmain.htm>
- North Carolina State Board of Education. (2000, December). *"Testing started with the ABCs" and other myths about testing and accountability in North Carolina*. Raleigh, NC: Author. Retrieved August 17, 2002, from <http://www.ncpublicschools.org/parents/myths.html>
- North Carolina State Board of Education. (2001, October). *A report on the impact of student accountability standards for grade 5 2000-2001*. Raleigh, NC: Author.
- Perlman, H. (2002, April 25). *Analysis finds 76 percent of class of 2003 meets state MCAS requirement*. Malden, MA: Massachusetts Department of Education. Retrieved May 1, 2002, from <http://www.doe.mass.edu/news/news.asp?id=595>
- Quality counts 2002: Building blocks for success [Special Report]. (2002, January 7). Bethesda, MD: Education Week.
- Shepard, L. (1991). When does assessment and diagnosis turn into sorting and segregation? In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 279-298). New York: Teachers College Press.
- Shepard, L., & Smith, M. (Eds.). (1989). *Flunking grades: Research and policies on retention*. London: Falmer Press.
- Texas Education Agency. (2000, November). *Texas Administrative Code, (TAC), Title 19, Part II, Chapter 101. Assessment*. Retrieved from <http://www.tea.state.tx.us/rules/tac/chapter101/ch101a.html#101.7>
- U.S. Department of Commerce, U.S. Department of Education, U.S. Department of Labor, National Institute of Literacy, & the Small Business Administration. (1999, January). *21st century skills for 21st century jobs*. Washington, DC: U.S. Department of Labor. Retrieved May 4, 2002, from <http://www.usworkforce.org/resources/pdf/skillsreport.pdf>
- U.S. Department of Labor, Bureau of Labor Statistics. (n.d.). *Working in the 21st century*. Washington, DC: U.S. Department of Labor. Retrieved May 7, 2002, from <http://www.bls.gov/opub/working/home.htm>