

## *Testing for a New Era*



*Furity.org*

*Marisa Bober  
Jane Cassie  
Cristin Datch  
Julie Tucker*

*A100: Introduction to Education Policy  
Professor Jal Mehta  
20 November 2009*

In the United States, the use of educational testing as a means of improving student learning has evolved over time. From states' efforts to hold students accountable, to the federal government's broad No Child Left Behind legislation, educational testing policy is omnipresent. How it will be used in the future to improve student learning is uncertain, but its application to teacher accountability is promising.

### ***HISTORY***

In order to understand the use of educational testing in the United States and its implementation and consequences, we need to look back at the legislation and social climate that preceded No Child Left Behind (NCLB). Before the Johnson administration, education was funded and controlled exclusively on state and local levels. Johnson, winning in 1964 by a wide majority and finding two-thirds Democratic majorities in both the House and Senate, was able to pass the Elementary and Secondary Education Act (ESEA) of 1965 (Davies, 2007). Although *Brown v. Board of Education* in 1954 involved the federal government in making decisions for schools with respect to desegregation (347 U.S. 483 (1954)), the ESEA was the first real foray into federal funding of education.

The ESEA was seen as part of the War on Poverty, providing Title I funds based on a formula that considered both the number of poor students in a district and the amount the district currently spent on each student. Because of this formula, over ninety percent of districts received some Title I funds. The ESEA was able to gain support not only because of its widespread distribution of funds, but also because funds followed students, not schools, even if they attended religious schools. Because spending was tied to students, Congress wanted to ensure that the funding it provided was actually being used for the intended populations. Thus began the first foray into school accountability for federal spending, which led to an over-emphasis on pull-out

programs not because of their effectiveness, but because of the ease of tying them to specific Title I-eligible students (Martin West, personal communication, Sept. 4, 2009).

The standards-based reform movement in American education was ignited by Secretary of Education Terrell Bell's 1983 seminal commissioned report on education, *A Nation at Risk* (Rudalevige, 2003). At the same time when American Federation of Teachers (AFT) president Al Shanker was calling for reform (Perlstein, 2007), *Nation* reported that America's students were falling dangerously behind in a world economy (*Nation at Risk*, 1983). However, President Reagan chose not to heed these warnings about America's K-12 education system.

Yet, throughout the 1980s and well into the 1990s, a national movement of standards-based reform, and the assessment projects necessary to measure those standards, was building steam. Only three years after *A Nation at Risk* came *A Nation Prepared*, the 1986 Carnegie Report calling for teacher standards nationwide and, importantly, tying those standards to student performance at the school level as a means toward equal and quality education (*A Nation Prepared*, 1986). Marc Tucker, who authored the report, headed the Commission in the Skills of the American Workforce,<sup>1</sup> which pushed for standards-based high school certificates indicating that students had met the high national standards that would prepare them to thrive in the global economy (*America's Choice*, 1990). The New Standards Project, which was “unquestionably the most ambitious” of these projects, was a coalition of states and districts trying to create assessments that would guide meaningful instruction (Spalding, 2000, p. 579). Co-founder Lauren Resnick, along with Tucker, summed up the tone of educational discourse at the time when she said in 1993, “the development of standards and assessments is a critical piece of reforming the entire educational system so that it is... driven by higher standards” (O’Neil, 1993,

---

<sup>1</sup> He now heads The *New* Commission on the American Workforce (emphasis in the original), which is still pushing for Board Exams after high school as an integral step to improving the American educational to the level needed for us to compete globally (NCEE, 2009).

p.17).

President George H.W. Bush, differing from his Republican predecessor and embracing the tone of education rhetoric in the nation, tried to shift the federal education discussion from access to quality. In 1990, the White House worked with governors including Bill Clinton to pass “Goals 2000,” an ambitious but toothless initiative to vastly increase American student competency by the year 2000 (Perlstein, 2007) by encouraging “states to develop process-related standards” (Fuhrman, 2004, p.134).

President Clinton’s 1994 reauthorization of the ESEA, the Improving America's Schools Act (IASA), focused on higher standards and included an increased strive toward accountability for meeting the standards that grew out of recent reforms. “The act said federal funding must be used to teach higher-level skills, and that children benefiting from Title I had to take standards-based tests three times during their schooling” (Perlstein, 2007, p.30), and state assistance and sanctions were contingent upon student scores. IASA required states to create content and performance standards in order to receive Title I funding, and emphasized “continuous and substantial yearly improvement” (P.L., in Rudalevige, 2003, p.29). Much like Goals 2000, IASA did not have much practical effect because states could set their own standards for improvement and had excessive time to comply with the law (Perlstein, p.30).

When President George W. Bush took office, he wasted no time announcing his education policy, sending his proposal to Congress on his second day in office (Perlstein, 2007, p.31). His proposal was centered on the importance of annual testing, with growth expected by each school and certain groups of students within each school every year. Many elements of his proposal, including yearly progress and required performance standards, can be seen in the past incarnation of the ESEA (Rudalevige, 2003, p.31).<sup>2</sup> The goal was to tie federal funds to

---

<sup>2</sup> Contentions along partisan lines over education policy from 1998-1999 led to a failure to reauthorize ESEA in

accountability measures in order to drive all schools and students each year toward meeting or exceeding a “proficient” standard. Bush considered accountability to be the “cornerstone of reform,” which was a change from past ESEA reauthorizations (Rudalevige, p.24). The President's plan, the 2001 reincarnation of ESEA entitled No Child Left Behind, was passed quickly with ninety percent support, although the process provided little time for feedback from educational groups (Perlstein, 2007, p.31) and the framing of the bill made opposing it a political difficulty (p.33). In the process of seeking bipartisan support, “accountability” was used in a variety of ways depending on its audience, yielding an ESEA that was both ambitious and ambiguous (Rudalevige, p. 24).

Though states can choose to reject the requirements of NCLB if they also reject its funds, no state has made this choice. Thus, they are compelled to meet proxies for teacher effectiveness. Firstly, NCLB calls for quality teachers in every classroom (NCLB, 2002), and requires teachers to prove their competency in their field of teaching through some state-defined criteria like focus of masters' degree or certification test (Perlstein, 2007, p. 32). States are also required to introduce annual testing for elementary- and middle-school students, and high school mathematics and language arts tests, targeted to state-set academic standards. States must define what constitutes the “proficiency” that increased percentages of their students must meet each year (Perlstein, p. 32). This improvement is called Adequate Yearly Progress (AYP), and measures each cohort of students within a school against the cohort from the previous year.<sup>3</sup>

The standard MCAS test was implemented in 1998 (prior to NCLB) as part of a series of reforms that essentially began with the Massachusetts Education Reform Act of 1993. This

---

1999. The proposals that came out of this process, including the House Republican's “Straight A's” bill, the Students Results Act, and Senate Democrat's “Three R's” amendment, contained language on many of the topics in NCLB (Rudalevige, 2003, p. 36).

<sup>3</sup> AYP must be shown not only for the grade as a whole, but also for certain subsets of the grade including minority students, Limited English Proficiency students, and students with disabilities (Rudalevige, 2003, p. 26).

legislation is said to have “turned around” Massachusetts’ education system, which had previously performed poorly on the National Assessment of Educational Progress (NAEP) on a consistent basis.<sup>4</sup> At this time, Massachusetts set state curricular standards and invested a great deal of new funds into education.

MCAS tests are given in language arts, math, science and technology, and history and social studies to all public school students in grades three through eight and grade ten, including those with disabilities or limited English proficiency (LEP). Each test is rooted in the curriculum framework for the state and its learning standards (MA Dept. of ESE, 2009).

Massachusetts is constantly reevaluating the MCAS. For example, after the first administration of the tenth grade math MCAS in 1998, the math test content was altered to remove coverage of trigonometry, reflecting the fact that many students in Massachusetts do not study trigonometry until pre-calculus in eleventh grade. Most recently, in early 2009, the Massachusetts Department of Education implemented four changes to the implementation and use of the MCAS: accelerated reporting of test results, shortened testing time for grades three and eight, release of a smaller percentage of test items (versus 100%), and use of a growth model to measure change between years (MA Dept. of ESE, 2009). The MCAS is used not only to meet national requirements of school accountability, but also at the state level to hold students accountable for their performance.

### ***POLICY RESEARCH***

In the U.S., policymakers and school administrators have long been utilizing “achievement tests as tools to hold students and, by extension, their teachers accountable for performance” (Koretz & Hamilton, 2006, p.533). Under NCLB, each state is required to establish its own definition of AYP (U.S. Department of Education, 2009). After two

---

<sup>4</sup> Much of the information on MCAS policy comes from an interview with Jesse Dixon, Special Assistant to Deputy Commissioner Karla Baehr at the Department of Elementary and Secondary Education, on 10/27/09.

consecutive years of failing to meet AYP requirements, a school is “identified for improvement” (U.S. Dept. of Ed., 2009, NP) and must notify its students’ parents of their right to send their children to another school. After three consecutive years, the school is put into the “corrective action” stage, wherein its local education agency (LEA) is required to take action for improvement; after four consecutive years, the school enters into the “restructuring” stage, wherein its LEA is required to choose from five restructuring plans. After five consecutive years of failing to meet AYP, the LEA is then required to implement its plan to restructure. At present, approximately 3,000 public schools in the U.S. are in the process of restructuring (Koretz & Hamilton, p. 540).

Increasingly, accountability standards have dictated curriculum and instruction, leading teachers to coach students on the content of their state’s test. This type of coaching often leads to score inflation. Professor Daniel Koretz asserts that score inflation occurs “when coaching generates gains that are limited to a specific test...and that do not generalize well to other tests of the same domain or to performance in real life” (Koretz, *Measuring Up*, 2008, p.255). The lack of actual knowledge that such score inflation can obscure is a significant negative effect of using large-scale high-stakes testing as the primary measure of student learning.

The three stated purposes of the MCAS are to: “inform/improve curriculum and instruction;” “to evaluate student, school, and district performance according to Curriculum Framework content standards and MCAS performance standards;” and “to determine eligibility for high school Competency Determination” (MA Dept. of ESE, 2009, NP). The above purposes indicate that the MCAS is a high-stakes test not only at the school level (with regards to making AYP) but also at the student level, as one must pass the tenth grade MCAS to graduate. Perhaps more telling, though, is its stated goal of informing curricula. This indicates that Massachusetts

lawmakers intended for the MCAS to dramatically impact curriculum in Massachusetts public schools, which it has.

Prior to 2008, the Office of Educational Quality and Accountability (EQA) had “served as an independent mechanism to verify the efforts of schools and school districts to promote a higher level of academic achievement by students” (MA Dept. of ESE, 2009, NP). In 2008, the Massachusetts State Legislature passed legislation that made a drastic change to Massachusetts’ accountability policies by dissolving the EQA and moving accountability back into the hands of the Massachusetts Department of Elementary and Secondary Education (MA DOE). As of June 2009, approximately 53% of Massachusetts schools were failing to make AYP, according to newly implemented state standards (MA Dept. of ESE). The MA DOE argues that this is because it has set extremely high standards for its schools (Jesse Dixon, personal communication, October 27, 2009), which is supported by Massachusetts’ high NAEP scores (Martin West, personal communication, Nov. 13, 2009).

The 2008 legislation that moved accountability into the hands of the MA DOE now allows for the state for the takeover, or “co-governance,” of a school as a consequence for failing to make AYP for five consecutive years. This “accountability system redesign” (MA Dept. of ESE, 2009, NP) will potentially see a greater number of schools shut down. As a result, the consequences of test scores increase, which puts more pressure on schools to emphasize the test over other student learning goals (Koretz, *Measuring Up*, 2008, p. 237).

Struggling schools in Massachusetts are indeed narrowing their curricula to focus on MCAS-tested content. Two-hundred fifty have had to abide by “Essential Conditions,” which require extra time for reading and math (MA Dept. of ESE, 2009, NP). Unfortunately, the type of narrowed instruction mandated by these conditions leads educators to pare down on subjects



elsewhere (Perlstein, 2007, p. 122). Massachusetts sees the means as worth the ends, as displayed by increasing MCAS scores. The 2009 Summary of State Results reports that “between 2008 and 2009, the percentage of students scoring *Proficient* or higher improved by two or more percentage points on four of the seven grade-level tests in English Language Arts (ELA), three of seven grade-level tests in Mathematics, and one of three grade-level tests in Science and Technology/Engineering” (MA Dept. of ESE, NP).

Additionally, the high-stakes aspect of these tests encourages states to lower their standards, leaving students unprepared for postsecondary education and/or the workplace. This lowering of standards is most visible when state scores are compared to NAEP scores. Chairman of the House Committee on Education and Labor, Congressman George Miller, acknowledged a recent study published by the National Center for Education Statistics that found that “most states are below or only meeting *basic* standards set by NAEP” (Miller, October 29, 2009, NP).

Evidence of coaching or teaching to the test is also apparent. Since the implementation of NCLB, NAEP scores in math and reading have not improved significantly even though many schools have focused on these areas (Ravitch & Chubb, 2009). Even though state scores might purport that gains in learning have been made, the validity of those inferences is undermined by teaching to the test and coaching. While Tennessee claimed that ninety percent of its fourth graders were proficient in reading, the NAEP indicated that only twenty-six percent were; similarly, North Carolina claimed that eighty-six percent of its fourth graders were proficient in reading, while the NAEP claimed that only twenty-eight percent were (Ravitch & Chubb, 2009). In 2007, the MCAS classified fifty-six percent of fourth graders and seventy-five percent of eighth graders as proficient in reading, while the NAEP labeled forty-nine percent and forty-three percent as proficient, respectively (MA Dept. of ESE, 2009). These discrepancies in test

scores suggest that students are not learning broadly. Clearly, many students pass the state test, but perhaps this is only because they are coached solely on the material on the state test; thus, how much better are schools really doing when so many students still score poorly on the National Assessment?

NCLB was established to provide quality education to all American students. It is often seen as a civil rights act because it aimed to reduce the achievement gap between students by ensuring that all students were well-served by schools—in essence, to leave no child behind (Darling-Hammond, 2007). Specifically, NCLB strived to provide better educational opportunities for racial minorities, students with disabilities, Limited English Proficiency (LEP), and low-income students by increasing their performance to a level of “proficiency.” It did this by requiring AYP for each of these individual groups instead of only entire schools.

It is no secret that wide disparities exist among schools. Some wealthy public school districts spend \$30,000 per student whereas some poorer districts spend only \$3,000 per student (Darling-Hammond, 2007). Disparities also exist in terms of school resources; in some high-poverty schools, textbooks are scarce and outdated and the heating system is nonfunctional (Darling-Hammond, 2007). Thus, students in urban, racially-segregated areas typically receive a substandard education (in comparison to their more affluent peers) due to a lack of resources (Payne, 2008) and dearth of quality teachers (Johnson et. al., 2004). Similarly, gaps exist between different subgroups of students, as on average, whites outperform blacks and Latinos, LEP students lag behind native-speakers, students with disabilities score lower than other students, and low-income students score lower than middle-class and affluent students (MA Dept. of ESE, 2009). NCLB aimed to rectify this social situation.

The efficacy of NCLB in narrowing the achievement gap is questionable. In

Massachusetts, MCAS scores between 2008 and 2009 reveal that the percentage of “proficient” or “advanced” students increased by at least two percentage points on many of the grade level tests (MA Dept. of ESE, 2009). While Grade 10 Math MCAS scores have continuously risen since 2002, from 2008 to 2009 the gap between the percentage of proficient white and black students did not narrow, nor did the gap between LEP students and other students, or the gap between low-income students and all other students (MA Dept. of ESE, 2009). While most subgroups have continuously improved their scale scores by a few points and more students are meeting proficiency; the gap between subgroups is still substantial. This is evidenced by the fact that while 85 percent of white Grade 10 students were proficient in math, only 56 percent of their Hispanic peers were (MA Dept. of ESE, 2009). Not only have the gaps persisted, but Massachusetts’ high school dropout rates have increased, particularly for African American and Hispanic students (MA Dept. of ESE, 2009). Ironically, the largest increases in test scores tend to appear in schools that have the highest dropout rates and high percentages of ninth grade students who were retained or went “missing” (Wheelock, 2003, cited in Darling-Hammond, 2007). Thus, while NCLB aimed to narrow the achievement gap, the data do not suggest this is happening (Darling-Hammond, 2007).

Massachusetts is not the only state struggling to improve the performance of its students. Rather than trying to boost the performance of low-scoring students, Texas schools have a tendency to make low-performing students disappear from the testing data (Darling-Hammond, 2007). Similarly, evidence suggests that New York has increased its scores by pushing out weaker students who probably would not pass the state’s graduation exam (Darling-Hammond, 2007). To prevent weak students (who would likely not pass the test) from ruining a school’s chance of meeting AYP, New York schools retain many ninth grade students or suspend low-

performing students on days before high-stakes exams (Amrein & Berliner, 2003). Thus, improving test scores does not necessarily indicate improved student learning. In fact, many students may actually be learning less, illustrating the inefficacy of NCLB (Hursh, 2007). Even those students being tested may be learning less than the tests indicate, as demonstrated by the disparity in state and NAEP scores shown above.

Nationally, policymakers and teachers alike have found that high-stakes tests in the NCLB era is narrowing curriculum. As one former principal and district administrator from Vermont aptly stated, this is because “people will teach to what we hold them accountable [SIC].”(Rebecca Holcombe, personal communication, November 16, 2009). As a result of NCLB, many changes within the classroom have occurred.

- The curriculum has been narrowed, teaching has become more teacher-centered, and information has become more fragmented (Au, 2007).
  - Because learning has become narrowly aligned to the test, students receive a fragmented education, with discrete pieces of information disconnected from a solid knowledge base (Ellis, 2008).
  - Perlstein’s (2007) work reveals the diluted education students receive based on a rigid definition of achievement, relegating many low-income students to an education that involve shrinking the breadth of the curriculum and ignoring certain subjects until after the high-stakes exam.
  - Research suggests that the implementation of high stakes tests has prompted teachers, especially in urban districts, to teach only the material that appears on exam (Amrein & Berliner, 2003).
- The U.S. Department of Education has reported that from 1988 to 2004 the time spent on history and social studies in elementary school declined by 22 percent while reading and English instructional time increased, but it is hypothesized that this time is spent primarily on basic reading skills (Hess, 2008).
- Disturbingly, a high school math teacher in an urban charter school noted that since the MCAS only tests content up to Grade 10, the school’s Grade 11 and 12 math program is very weak; thus by focusing on a narrow goal—MCAS scores—other areas have been neglected (Anonymous personal communication, October 16, 2009).
- Because Texas students must pass a writing component of the assessment, teachers tell students how to write formulaic responses that will earn full credit, but this clearly dilutes the breadth and depth of writing (Hursh, 2007).
- Schools now spend a lot of their budgets on test-prep materials rather than on dynamic materials or books, clearly illustrating that the focus on meeting AYP has reduced the

breadth of other opportunities available to children (Hursh, 2007).

It is difficult to imagine how learning can improve with the aforementioned decline in the richness of opportunities available to students.

Under its current structure and implementation, NCLB hurts the students it aimed to benefit (Darling-Hammond, 2007). While President Bush claimed standards were needed to eradicate “the soft bigotry of low expectations,” Perlstein (2007) powerfully notes that “to condemn [students] to a rudimentary education in the name of improvement is bigotry, too” (p. 136). She criticizes NCLB for depriving students of a real learning experience, as children are forced to regurgitate fragmented facts. Although NCLB had well-meaning purposes, it has created an educational system in which achievement rests solely on *one* scale score on *one* test on *one* day that purports to measure proficiency on everything learned. Thus, the implementation of NCLB policy has not effectively achieved its goal of narrowing the achievement gap and leaving no child behind. In looking at the trends, more students have become “proficient” or “advanced” since 2002, but given evidence of push-outs, retention, curriculum-narrowing, and coaching, are these data meaningful?

The American Recovery and Reinvestment Act of 2009 released emergency funds to states for education and set aside \$4.35 billion for the Race to the Top competitive grants program. Although the Department of Education has issued discretionary grants in the past, Race to the Top is the “largest competitive education grant program in U.S. history.”(U.S. Dept. of Ed., 2009, NP). This fund has been “designed to provide incentives to states to implement large-scale, system-changing reforms that result in improved student achievement, narrowed achievement gaps, and increased graduation and college enrollment rates”(Miller, September 24, 2009, NP).

As for the future of NCLB and its accountability measures, policymakers seem uncertain as to what the next reauthorization of the ESEA may look like. House Chairman Miller has indicated that federal policymakers must “improve the law in reasonable and commonsense ways that reflect what [they’ve] learned since the last reauthorization and that build on the important reforms outlined in the Race to the Top Fund” (Miller, September 24, 2009, NP). With regards to the accountability system set up by NCLB, Miller stated that policymakers “must ensure that all students are held to rigorous standards... [and] must measure student achievement more accurately and fairly” (NP). Whether or not these high standards will continue to be measured using high-stakes tests as tools remains to be seen.

### ***CONTEMPORARY POLITICS***

The most widespread example of testing for accountability today is the testing by each state to satisfy the NCLB requirements. NCLB started with bipartisan support from both major political parties; however, since that day the tone has shifted. Today, there are essentially no groups that fully support NCLB in its current state; however, different groups have taken various views on different aspects of the NCLB. It is helpful to understand that accountability can be imposed on three general levels: school accountability, teacher accountability, and student accountability. School accountability includes punishing schools for not meeting requirements such as AYP. Teacher accountability includes using student test scores or other measurements to reward or punish teachers. Student accountability includes using test scores as a promotional or graduation requirement. NCLB, as written, only holds school accountable for AYP. Some states, such as Massachusetts, have used these tests for student accountability by using tests as a graduation requirement. Each group described below has been fit into this framework for purposes of comparison.

Ted Kennedy, the staunchest Democratic supporter of NCLB in 2001, was a proponent of

school and teacher accountability; however, after the law's enactment he changed his mind about the way to hold these parties accountable. Early in 2008, Kennedy penned an essay in the *Washington Post* detailing the necessary changes he thought were needed to repair the accountability requirements of NCLB. His plans included using a wider scope of measurements, in addition to using test scores, to evaluate the effectiveness of schools (Kennedy, 2008). Additionally, he wanted to remove the punitive score threshold requirements and instead reward schools for incremental improvements. Also, he was concerned about rewarding teachers for teaching to the test, so he wanted to promote innovative teaching rather than just rewarding high scoring teachers (Kennedy, 2008).

Groups such as the national Parent Teacher Association (PTA) and the National Association for the Advancement of Colored People (NAACP) support using tests along with other measurements to hold schools accountable. However, both the PTA and the NAACP have expressed an interest in moving away from measuring AYP with a linear model. In a written response to President Bush's 2007 "State of the Union" address, the NAACP stated that AYP requirements are too punitive. It recommended that we judge schools based on how much scores have grown (NAACP, 2007). A statement on the PTA's website endorses using growth models to measure schools (PTA, n.d.). The PTA and NAACP have additional suggestions for school accountability beyond testing. For example, the NAACP recommends that NCLB include a standardized formula for calculating dropout rates and then hold schools accountable to these rates (NAACP, 2007). The PTA's chief area of concern, and great disappointment in NCLB, is that schools should be held accountable for including parents in their child's education (PTA, n.d.).

The Education Trust is one of the strongest supporters of holding schools and teachers

accountable through student testing. The Education Trust has stated that the best accountability systems for making real change in schools are those that focus on results alone (Education Trust Statement, 2007). It believes that we should neither look at the processes that go into education nor should we try to collect data on as many indicators as we can (Testimony of the Education Trust, 2004). In testimony given in 2007 to California's Department of Education, the Education Trust urged Californians to implement every aspect of NCLB including NCLB's aggressive goal of 100% proficiency by 2014 (Ali, 2007). In addition to AYP accountability for schools, the Education Trust also supports individual teacher accountability. In the Education Trust's Winter 2004 publication of *Thinking K-16*, Kevin Carey (2004) addresses and promotes the use of value-added models to determine which teachers are effective and which teachers should be let go.

The Education Equality Project (EEP) supports school, teacher, and student accountability. In the EEP's *Paper Series on Accountability*, The EEP proposes using value-added measures of student growth in the core academic subjects with accountability on all three levels of our framework. EEP also proposes offering incentives for improvement in order to get schools and teachers to more willingly accept performance assessments. If a school fails to show growth, it should be shut down rather than fixed; however, if a school performs well, then teachers and principals should be financially rewarded. In order to reduce some of the punitive measures of NCLB, the EEP suggests that schools still be considered as meeting AYP even if they narrowly miss AYP due to a single subgroup. Additionally, the EEP proposes giving monetary rewards to individual students who show significant growth (Education Equality Project, 2009).

The two national teacher's unions, the AFT and the National Education Association



(NEA) disagree with multiple provisions of NCLB. The official statement of the NEA is that NCLB includes “draconian penalties for schools that cannot meet a series of one-size-fits-all standards” (NEA, n.d., NP). The AFT concurs with this view and further stated in a letter to Secretary of Education, Arne Duncan, that NCLB is too narrow and not research-based (Weingarten, 2009, September 25). On April 29, 2009, Randi Weingarten, president of the AFT, testified before the House Committee on Education and Labor. In her testimony, she advocated for the creation of national standards and accountability in order to reduce the variability in standards across states. She also stated that accountability should not be used to blame schools; rather, it should be used as a way to help fix schools (Weingarten, 2009, April 29). Weingarten is also open to using test scores for teacher accountability by using the scores as a part of calculating teacher compensation (Weingarten, 2009).

In the NEA’s 2008 “Great Public Schools for Every Student by 2020,” the NEA asserts that school reform should be accomplished through multiple measures of school and teacher accountability (NEA, 2008). In the *Great Public Schools* report, the NEA advocates for schools to be rated on their quality of education as measured by the level of their “rich and comprehensive curriculum” as well as the assessment of student learning (NEA, 2008, NP). As a way to improve the quality of teaching, the NEA also proposes that the federal government play a role in raising the bar for entry into the teaching profession (NEA, 2008).

At the *Education Stakeholders Forum* on September 24, 2009, the U.S. Secretary of Education, Arne Duncan, articulated the problems he saw with accountability in NCLB. At this meeting Secretary Duncan stated that he will form his proposal for the next reauthorization of the ESEA only after he completes a series of public conversations around the country to gather ideas and get input from education stakeholders. The accountability items he specifically refers to

include better tests that measure the right things, recognizing schools for their growth rather than punishing them for missing AYP, and measuring students by growth rather than absolute test scores (Duncan, 2009, September 24). Secretary Duncan, along with President Obama, is also a strong proponent of merit pay for teachers (Bruce, 2009). He believes that student test scores should play a prominent role in teacher accountability for merit pay, though he also acknowledges that single test scores alone are not enough to capture evidence of effective teachers (Bruce, 2009). This belief in holding teachers accountable is evidenced by Race to the Top's requirement that states allow student and teacher data to be linked.

One of the barriers to improving schools through high-stakes testing is that test scores alone cannot fully measure the quality of those schools; unfortunately, there is no known better alternative. At the same time, value-added models are not an accurate measurement of teacher quality, either (D. Koretz, personal communication, November 16, 2009). Race to the Top, with its allowance for states to try new methods for improving schools, and the next reauthorization of the ESEA are our opportunities for change.

### ***POLICY ANALYSIS***

In an effort to improve student learning, student assessments have been used by a number of actors to drive a variety of improvements. Once the federal government became involved in education policy, it strove to tie its funding to closing the achievement gap. Schools were only held accountable with regards to the funding allocations. Thus, student assessments began being used as a measure of student learning. The federal government, most recently in No Child Left Behind, has used student test scores to hold schools accountable for making AYP toward the goal of 100 percent student proficiency. Even prior to NCLB, many state governments embraced the usefulness of test scores, using them for student level accountability; in Massachusetts, for example, students are held accountable for MCAS scores because their graduation hangs in the

balance.

More recently, focus on the national and state levels has turned to using student test scores as a means of holding teachers accountable. This change has come about perhaps because the past methods have been so ineffective at improving student learning. Holding schools accountable is wise in theory, because schools are a collective unit in ways that districts or teachers are not. Actors within schools have the ability to work collaboratively, which some studies have shown to improve teaching and student learning (Kapadia et. al, 2007). Schools also share an administration, so accountability at the school level includes an indirect assessment of the leadership and does not, for better or worse, assign the effects of the administration on student learning as the successes or failures of teachers or students. However, many measures indicate that it has failed to improve student learning, and numerous examples of its perverse consequences to student learning (Darling-Hammond, 2007; Perlstein, 2007) are rearing their ugly heads. Sam Houston of North Carolina's Blue Ribbon Commission on Testing and Accountability said, "We're testing more but we're not seeing the results. We're not seeing graduation rates increasing. We're not seeing remediation rates decreasing. Somewhere along the way testing isn't aligning with excellence" (Ravitch & Chubb, 2009, p. 51).

Holding students accountable has a sound justification, too. Systems of accountability create incentives for behavior. If students are held accountable for a certain set of knowledge and skills, and are made to prove their minimal ability in those areas or be denied a high school diploma, they will have an added incentive to gain those skills (Sykes, 1983). However, this policy presumes that students have the resources and abilities to make the improvements necessary, and it is clear that many of them do not.

The idea behind teacher accountability is fairly straight-forward. We know that the

quality of teachers in the classroom makes a measurable and dramatic impact on student learning (Johnson et. al., 2008). Some studies suggest that it makes a bigger impact than any other factors we have measured (Goldhaber, in Birkeland & Curtis, 2002). Therefore, improving teachers may very well be the best way to improve student learning. Even if improving schools and improving individual student performance are wise roads to reform, improving teacher quality is a way of moving toward both of those. In order to improve the quality of teachers in our classrooms, we need a way to measure teacher effectiveness, as well as a method to build upon it once measured.

Teachers themselves tend to define their effectiveness in terms of “student achievement and motivation” (Johnson & Birkeland, 2002, p.105). Unfortunately, assessments are not infallible. Tests are only helpful information in as much as their results are indicative of the broader domain they are testing. Thus, in a policy focused on teacher accountability, tests are only helpful if the inferences we draw from them about teacher quality are valid. The best policy, then, would do two things. Firstly, it would strive to make the inferences we draw about teacher quality as valid as possible. One of the best ways to do that is to include more than one measure of teacher quality. This could be two different student assessments, or a student assessment combined with another measure (like co-teacher observations).

Secondly, a desirable policy would be able to assess the effectiveness of its assessments. Policies can accomplish this by evaluating the results of their assessment-based evaluation of teacher effectiveness by comparing them to another measure of student learning. This type of evaluation is not problem-free, because there could be inaccuracies in the comparison measurement. But it is an indication of the accuracy of the chosen measure of teacher quality. This comparison loses most of its value if the comparative measure can be gamed in the same ways as the desired measurement (Loeb and Rieninger, 2004). However, the caveat is that

comparative assessments must be low-stakes for teachers. The following hypothetical example accomplishes these twin goals of effective teacher quality measures: maximizing the validity of assessment inferences and evaluating the measurement through comparison.

*The state wants to improve student learning, and has decided to use student assessments to evaluate teacher quality. The student assessment is given at the start of the school year, with a comparable assessment given at the end of the school year. Each one takes a full day. It tests a subsection of a list of content knowledge and student skills in a variety of different ways. Teachers will be provided with this list at the start of the year. Student scores at the end of the year will be measured against other students' scores in the state in order to show the percentile rank on this test score. That percentile rank will be compared with the student's percentile rank at the start of the year. The state will look at the average change in percentile rank for each teacher's classroom of students. The state will also look at teacher evaluations made by lesson observers. Those teachers that are identified as least effective by this combination of measures will receive professional development training and resources to improve their teaching so that it may become more like the teaching of the most effective teachers. Every two years, students will be given an assessment styled after the NAEP, known to be a fairly accurate indication of proficiency (Koretz, Measuring Up 2008). If our assessments of teacher quality are correct and our interventions are effective, student learning should be improving. If it is not improving, our policy is identifying the wrong teachers as effective, and needs to be adjusted.*

It is important for us to see that measurement alone does not quality teachers make. Even though we have these incentives in place, it is not like people automatically know what to do or how to fix the problem (Elmore, 2004). This element is missing from a punitive system like No Child Left Behind, which seems to presume that schools and teachers could help students improve faster if only they were willing and correctly incentivized. This has turned out not to be true. The assessments, as in the example above, must be combined with some kind of professional development system that provides ineffective teachers with the skills and resources necessary to improve.

How can the federal government encourage models like this? By funding them. These models necessarily require some evolution over time, so it would be wise to test them in smaller subsets of schools and then apply those assessments and methods that are shown to work.

Testing expert Daniel Koretz suggests that one effective solution might be to use schools as

laboratories on a small scale, by trying these initiatives in a small subset of schools or districts before implementing them on the whole (personal communication, November 17, 2009). These evaluations of teacher quality should be low-stakes at first, until we know how accurate they are. The combinations of assessment and remedial professional development that work can then be applied at scale.

Fortunately, the current administration's Race to the Top initiative seems like an effort to do just this. President Obama and Secretary Duncan have spoken about the use of assessments containing more complex tasks as potentially effective ways for measuring student learning, and the use of value-added models as potentially effective ways of measuring how much of the student learning is attributable to the teacher. By incentivizing the creation of new student assessments, and new ways of using them to measure and improve teacher quality, this administration is moving in the right direction. The proposal also allows for flexibility among states to focus on whichever initiatives they find to produce student learning, which the rigidity of No Child Left Behind failed to do (Darling-Hammond, 2007).

This policy is not without criticisms. Though Race to the Top does not explicitly require value-added models of assessing teacher effectiveness, the administration has suggested these models as favorable options. But value-added models are not perfect. For example, in value-added models, it is difficult to disaggregate the effects of the teacher's efficacy and the effects of external influences not attributable to the teacher's efforts (Koretz, *A Measured Approach*, 2008). Additionally, value-added measures do not decrease the inclination to inflate test scores, as there would be no guarantee that teachers would not narrow their curriculum to boost their own students' scores. Additionally, the general idea of measuring teachers, and potentially removing them based on those measurements, has been criticized by unions. The main criticism is that

those teachers serving certain groups of students will suffer unjust repercussions, although value-added assessments are one effort to isolate teacher effectiveness from other influences on student learning (Elmore et. al., 1996). It is important to understand that value-added models can be a huge asset for teachers and to frame them as a resource to the teaching community in order to gain buy-in from unions and their supporters. Teachers who are unable or unwilling to improve may be counseled out of the profession or denied tenure. However, the main goal of this policy is to improve the quality of teachers' instruction in our classrooms, thereby improving student learning. Those teachers whose students do not progress would not be punished; instead, they would be provided with the resources and training they need to improve. This policy helps us begin to isolate effective teaching practices so that they can become the shared best practices of a profession, thus equipping teachers with the tools to do their jobs more effectively.

### Works Cited

- A Nation Prepared: Teachers for the 21st Century*. (1986). Hyattsville, MD: Carnegie Forum on Education and the Economy.
- Ali, R. (2007). Testimony on the reauthorization of NCLB provided to the California department of education on January 2, 2007. Retrieved from <http://www.edtrust.org/sites/edtrust.civicaactions.net/files/publications/files/Testimony%20on%20the%20Reauthorization%20of%20NCLB%20Provided%20to%20the%20California%20Department%20of%20Education.pdf>.
- Amrein, A.L. & Berliner, D.C. (2003). The testing divide: New research on the intended and unintended impact of high-stakes testing. *Peer Review*, 5(2), 31-32.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Bersin, A. (2006). *A theory of action for high school reform*. New York: Carnegie Corporation of New York.
- Birkeland, S.E. & Curtis, R. (2006). *Ensuring the support and development of new teachers in the Boston Public Schools: A proposal to improve teacher quality and retention*. Boston, MA: Boston Public Schools.
- Brown vs. Board of Education of Topeka, 347 U.S. 483 (1954).
- Bruce, M. (2009, July 2). Duncan stresses merit pay to teachers union. *ABC News*. Retrieved from <http://abcnews.go.com/Politics/story?id=7991577&page=1>
- Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K-16*, 8(1), The Education Trust. Retrieved from <http://www2.edtrust.org/NR/rdonlyres/5704CBA6-CE12-46D0-A852-D2E2B4638885/0/Spring04.pdf>
- Committee on Education and Labor. (2009). Miller Press Release: New Report Shows Too Many States Weakening Education Standards. Retrieved on November 14, 2009, from <http://edlabor.house.gov/newsroom/2009/10/miller-new-report-shows-too-ma.shtml>
- Committee on Education and Labor. (2009). Chairman Miller Statement on Secretary Duncan's First Major Speech on the Future Reauthorization of ESEA. Retrieved on November 14, 2009, from <http://edlabor.house.gov/newsroom/2009/09/chairman-miller-statement-on-s-3.shtml#more>
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind.' *Race Ethnicity and Education*, 10(3), 245–260.



- Davies, G. (2007). *See government grow: Education politics from Johnson to Reagan*. Lawrence, Kansas: University Press of Kansas.
- Duncan, A. (2009, September 24). Speech at the monthly meeting of the Education Stakeholders Forum. Retrieved from <http://www.ed.gov/news/pressreleases/2009/09/09242009.html>
- Education Equality Project. (2009). Education Equality Project position paper series on Accountability: August 11, 2009. Retrieved from [http://eep.3cdn.net/45852972c23ac04c00\\_v0m6bhgvn.pdf](http://eep.3cdn.net/45852972c23ac04c00_v0m6bhgvn.pdf).
- Education Trust Statement (2007). Education Trust statement on FEA proposals for assessment and accountability. Retrieved from <http://www.edtrust.org/dc/press-room/press-release/education-trust-statement-on-fea-proposals-for-assessment-and-accountabi>
- Ellis, M. (2008). Leaving no child behind yet allowing none too far ahead: Ensuring (in)equity in mathematics education through the science of measurement and instruction. *Teachers College Record*, 110(6), 1330–1356.
- Elmore, R.F. (2004). *School reform from the inside out: Policy practice and performance*. Cambridge: Harvard Education Press.
- Elmore, R.F., Abelman, C.H., & Fuhrman, S. (1996). The new accountability in state education reform: From process to performance. In H.F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (65-98). Washington, D.C.: Brookings Institution Press. Federation for Children with Special Needs. (2004). Supplemental educational services under NCLB. Retrieved from <http://ppplace.org/publications/pointers/English/pppses.pdf>
- Fuhrman, S. (2004). Less than meets the eye: Standards, testing, and fear of federal control. In N. Epstein (Ed.), *Who's in Charge Here? The Tangled Web of School Governance and Policy* (131-163). Washington, D.C.: Brookings Institution Press.
- Goldhaber, D. (2002). Teacher quality and teacher pay structure: What do we know and what are the options? *Georgetown Public Policy Review*, 7(2), 81-94.
- Hess, F. (2008). Still at risk: What students don't know, even now. *Common Core*. Retrieved from [http://www.commoncore.org/docs/CCreport\\_stillatrisk.pdf](http://www.commoncore.org/docs/CCreport_stillatrisk.pdf)
- Hursh, D. (2007). Exacerbating inequality: The failed promise of the No Child Left Behind Act. *Race Ethnicity and Education*, 10(3), 295–308.
- Jennings, J., & Rentner, D. (2006). *Ten big effects of the No Child Left Behind Act on public schools*. Washington, D.C.: Center on Education Policy.
- Johnson, S.M. & Birkeland, S.E. (2008). Is fast-track preparation enough? It depends. In P.

- Grossman & S. Loeb (Eds.), *Taking stock: An examination of alternative certification* (101-128). Cambridge: Harvard Education Press.
- Johnson, S.M., Kardos, S.M., Kauffman, D., Lin, E., & Donaldson, M.L. (2004). The support gap: New teachers' early experiences in high-income and low-income schools. *Education Policy Analysis Archives*, 12(61), 1-24.
- Kapadia, K., Coca, V. & Easton, J. (2007). *Keeping new teachers: A first look at the influences of induction in the Chicago Public Schools*. Chicago Consortium for School Research. Retrieved from [http://ccsr.uchicago.edu/content/publications.php?pub\\_id=113](http://ccsr.uchicago.edu/content/publications.php?pub_id=113)
- Kennedy, T. (2008, January 7). How to fix 'No Child'. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-dyn/content/article/2008/01/06/AR2008010601828.html>
- Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge: Harvard University Press.
- Koretz, D. (2008). A Measured Approach: Value-added measures are a promising improvement, but no one measure can evaluate teacher performance. *The American Educator*, Fall 2008. Retrieved from: [http://www.aft.org/pubs-reports/american\\_educator/issues/fall2008/koretz.pdf](http://www.aft.org/pubs-reports/american_educator/issues/fall2008/koretz.pdf).
- Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.), (531-578). Westport, CT: American Council on Education/Praeger.
- Loeb, S., & Reininger, M. (2004). *Public policy and teacher labor markets: What we know and why it matters*. Retrieved from ERIC database. (ED485592)
- Massachusetts Department of Education. (2007). *Guide to Interpreting the Spring 2007 MCAS Reports for Schools and Districts*. Massachusetts Department of Education. Retrieved from [http://www.doemass.org/mcas/2007/interpretive\\_guides/full.pdf](http://www.doemass.org/mcas/2007/interpretive_guides/full.pdf)
- Massachusetts Department of Elementary and Secondary Education. (2009). *Board in Brief: The Massachusetts Board of Elementary and Secondary Education*. Retrieved from <http://www.doe.mass.edu/boe/bib/09/0623.html?printscreen=yes&>
- Massachusetts Department of Elementary and Secondary Education. (2009). *Spring 2009 MCAS Results: Summary of State Results*. Retrieved from <http://www.doe.mass.edu/mcas/2009/results/summary.pdf>
- Massachusetts Department of Elementary and Secondary Education. (n.d.) *Massachusetts Comprehensive Assessment System Overview*. Retrieved October 19, 2009, from <http://www.doe.mass.edu/mcas/overview.html>

- Massachusetts Department of Elementary and Secondary Education. (n.d.). *School and District Accountability and Assistance*. Retrieved October 21, 2009, from <http://www.doe.mass.edu/sda/review/district/EQA.html?section=overview>
- Massachusetts Department of Elementary and Secondary Education. (n.d.). National Assessment of Educational Progress: MCAS & NAEP Comparison. Retrieved October 21, 2009, from <http://www.doe.mass.edu/mcas/naep/compare.html>
- Mehrens, W. (1992). Using performance assessments for accountability. *Educational Measurement: Issues and Practice*, 11(1), 3-9.
- National Association for the Advancement of Colored People. (2007). The Unstated state of the union. *The NAACP Advocate*, 2(1). Retrieved from [http://www.naacp.org/pdfs/advocate/Advocate\\_JanFeb2007.pdf](http://www.naacp.org/pdfs/advocate/Advocate_JanFeb2007.pdf)
- National Center on Education and the Economy. (2007). *Tough choices or tough times: The report of the New Commission on the Skills of the American Workforce*. Retrieved from <http://www.skillscommission.org/executive.htm>
- National Center on Education and the Economy. (1990). *America's choice: High skills or low wages! The report of the Commission on the Skills of the American Workforce*. Retrieved from [http://www.skillscommission.org/pdf/High\\_SkillsLow\\_Wages.pdf](http://www.skillscommission.org/pdf/High_SkillsLow_Wages.pdf)
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, D.C.: U.S. Department of Education. Retrieved from <http://www.ed.gov/pubs/NatAtRisk/title.html>
- National Education Association. (n.d.). No Child Left Behind Act NCLB/(ESEA) [Web log post]. Retrieved November 19, 2009, from <http://www.nea.org/esea/>
- National Education Association. (2008). Great public schools for every student by 2020. Retrieved from <http://www.nea.org/assets/docs/GPS2020.pdf>
- O'Neil, J. (1993). On the New Standards Project: A conversation with Lauren Resnick and Warren Simmons. *Educational Leadership*, 50(5), 17-21.
- P.L. 103-382, Section 1111(b).
- Payne, C.M. (2008). *So much reform, so little change: The persistence of failure in urban schools*. Cambridge: Harvard Education Press.**
- Perlstein, L. (2007). *Tested: One American School Struggles to Make the Grade*. New York: Henry Holt Publishers.
- Parent Teacher Association. (n.d.). PTA position overview for the upcoming ESEA-NCLB reauthorization. Retrieved November 19, 2009, from <http://www.pta.org/NCLB->

## Summary-en.pdf

- Ravitch, D. (2003). An historic document. *Education Next*, 3(2), 33-38.
- Ravitch, D. & Chubb, J.E. (2009). The future of No Child Left Behind. *Education Next* 9(3), 48-56.
- Rudalevige, A. (2003). No Child Left Behind: Forging a congressional compromise. In P.E. Peterson and M.R. West (Eds.), *No Child Left Behind? The politics and practice of school accountability* (23-54). Washington, D.C.: Brookings Institution Press.
- Spalding, E. (2000). Performance assessment and the New Standard Project: A story of serendipitous success. *Phi Delta Kappan*, 81(19), 758-764.
- Sykes, G. (1983). Public policy and the problem of teacher quality: The need for screens and magnets. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (7-125). New York: Longman.
- U.S. Department of Education. (2009). Education department seeks advice from experts on ways to improve assessments. Retrieved from <http://www.ed.gov/news/pressreleases/2009/10/10202009.html>
- U.S. Department of Education. (2009). *Race to the Top Program: Preamble and Major Changes*. Washington, D.C.: U.S. Department of Education. Retrieved from <http://www.ed.gov/programs/racetothetop/major-changes.pdf>
- U.S. Department of Education. (2002). Key policy letters signed by the Education Secretary or Deputy Secretary. Retrieved from <http://www.ed.gov/policy/elsec/guid/secletter/020724.html>
- Weingarten, R. (2009, September 25). Letter to Arne Duncan. Retrieved from <http://www.pft400.org/AFT%20Comments%20on%20School%20Improvement.pdf>
- Weingarten, R. (2009, April 29). Testimony of Randi Weingarten, President, American Federation of Teachers, Before the House Committee on Education and Labor. Retrieved from <http://edlabor.house.gov/documents/111/pdf/testimony/20090429RandiWeingartenTestimony.pdf>
- Weingarten, R. (2009). What matters most: Public school entrepreneurs. [Web advertisement]. Retrieved from <http://www.aft.org/presscenter/speeches-columns/>

**List of Sources Interviewed:**

- Danny Allen, head of the ELA MCAS pull-out program at Chelsea High School in Chelsea, MA, 35-year veteran teacher and former head of the English Department at Chelsea High School.
- Jesse Dixon, Special Assistant to Deputy Commissioners Karla Baehr at the Massachusetts Department of Elementary and Secondary Education
- Rebecca Holcombe, HGSE Doctoral Student, Former Principal and Director of Academics for Rivendell Interstate School District for K-12 schools in Vermont
- Dan Koretz, Henry Lee Shattuck Professor of Education, Harvard Graduate School of Education
- An urban charter school math teacher and chair of math department (who wishes to remain anonymous)
- An urban charter high school principal (who wishes to remain anonymous)
- An urban public high school math teacher (who wishes to remain anonymous)
- An urban public high school English teacher (who wishes to remain anonymous)
- An urban public high school administrator (who wishes to remain anonymous)