

II. Testing - What's the Problem with 'No Child Left Untested'?

“The statewide test bombards the schools with pressure... In the second week of school we get things from the teacher like, “This was on the test last year so listen up”... The pressure restricts teachers from doing their job... They can't help struggling students fully understand the material because when the student starts to finally kind of get it, it's time to move on so they can get the entire curriculum taught. ... The test pressure just about kills some kids. I have never heard a student say, All this pressure from the test gets me fired up! More often, I see kids cracking. They start freaking out... the test is taking away the real meaning of school. Instead of learning new things and getting tools for life, the mission of the schools is becoming to do well on the test.”

— Seventh-grade student quoted in *Louisville Courier Journal*, February 10, 2003

“Nancy Baker, a reading teacher at Bristol Borough Junior-Senior High School, is discouraged by the emphasis on test-taking. Her students, who used to write reports on 11 books a year, now read only eight books because they have to prepare for the Pennsylvania System of School Assessment test, which relies on short-paragraph writing and multiple-choice questions.”

— “Battles Ahead over No Child Left Behind,” *Philadelphia Inquirer*, by Dale Mezzacappa, Toni Callas and Kellie Patrick, Nov. 18, 2003.

“Accountability should not rest on a single day, a single hour, a single testing situation. A North Carolina study found that 80 percent of teachers spent more than 20 percent of their time practicing for tests—that's not real learning. The single most important factor for improving [student achievement] is the teacher. It is tragic that a law designed to help students is actually driving teachers from the field.”

— Jill Morningstar, Children's Defense Fund

A. As Texas Education Goes, So Goes the Nation?

Texas, which gave us President George W. Bush and his Secretary of Education, Rod Paige, is often cited as the model for NCLB. On the strength of claims that he raised test scores, lowered the dropout rate and narrowed the race-based achievement gap, Rod Paige went from

Texas, which gave us President George W. Bush and his Secretary of Education, Rod Paige, is often cited as the model for NCLB.

superintendent of the Houston Independent School District (HISD) to the nation's top education official and pitch man for NCLB.

Over the past year, however, Paige's record in Texas has come under intense scrutiny. As a result, the premise that test-based reform will be the key to ensuring that no child really is left behind is now being questioned in prominent places like the front page of *The New York Times* and on CBS's *60 Minutes*.

The Texas Record

Since Texas's record was used to promote NCLB, it seems fair to look at the Texas experience and ask if this approach to education reform is likely to benefit or harm U.S. public schools. Here's the Texas record in sum:

- While racial score gaps closed on the now-replaced Texas Assessment of Academic Skills (TAAS), the gaps are as wide or wider on the new and more challenging Texas Assessment of Knowledge and Skills (TAKS) than they were when TAAS was first administered. Among high school students, the gap between white students and blacks and Hispanics is nearly 40 points in math and science. In 1994, the race-based gap averaged 30 points (Peabody, 2003). Gaps also remain quite large on national standardized tests like the Stanford Achievement Test (Schemo & Fessenden, 2003).
- A study by the RAND Corp. in 2000 found that while the gap in average scores between whites and students of color was decreasing according to TAAS results, it was increasing on the federal National Assessment of Educational Progress. The study's authors concluded the results "raise important questions about the generalizability of gains reported on a state's own assessment, and hence about the validity of claims regarding student achievement" (Klein *et al.*, 2000).
- The University of Texas has reported that the need for remediation had substantially increased since the advent of the TAAS graduation requirement, a need that could not be explained by the growth of the university or the decision to offer admission to the top 10 percent of all high school graduates (Haney, 2000).
- The dropout rate in Texas and in particular in its major cities remains very high. One study found that graduation rates in five Texas districts are among the two-dozen worst in the nation (Haney, 2001).

Since Texas's record was used to promote NCLB, it seems fair to look at the Texas experience and ask if this approach to education reform is likely to benefit or harm U.S. public schools.

Graduation rates in five Texas districts are among the two-dozen worst in the nation.

- Recently, charges of altering data in Houston to hide evidence of the district's high dropout rate have gained national prominence: While Houston graduated less than half its students in four years, the official dropout rate was only 1.5 percent (Schemo, 2003).
- A front-page *New York Times* story on July 30, 2003, confirms what has been found in many jurisdictions: Students are pushed out of school in order to boost test scores (Lewin and Medina, 2003).
- Another HISD technique to improve test results was to retain low-scoring students in ninth grade for years to avoid having them take the tenth grade test. Perla Arredondo finally dropped out after three years in ninth grade. "They used me and some other kids to make the school look better," she said, by holding students back to improve the school's tests scores. "It was all this three years in ninth grade. Because of the test they wouldn't let us move up" (Werner, 2003).
- William Bainbridge and Steven Sundre of the University of Dayton found that in the 1998-1999 school year, HISD had 18,221 seventh graders, but that two years later there were only 9,138 ninth graders, a middle school dropout rate of 53 percent. (Bracey, 2003).

Assembly-Line Education for Assembly-Line Jobs

A 2002 *Texas Observer* article by Jake Bernstein, "Test Case: Hard Lessons from the TAAS," examined the Texas TAAS legacy in some detail (Bernstein, 2002). The article concludes that "[TAAS] produces a class of students who will be perfect employees for a low-wage economy. They will lack training in critical thinking and be unprepared to find knowledge in the information age. It's not a good recipe for a vibrant democracy."

Referring to the classroom impact of TAAS on teaching, teacher Becky Mcadoo told Bernstein, "It became like an assembly-line education. Nothing mattered but the TAAS." The *Observer* reported federal data showing the teacher resignation rate in Texas climbed from 8.6 percent to 11.3 percent from 1997 to 2001.

"Under pressure from politicians, businessmen and administrators, school districts consistently inflate scores," the article concludes. "There are various ways to game the system." These include

"[TAAS] produces a class of students who will be perfect employees for a low-wage economy. They will lack training in critical thinking and be unprepared to find knowledge in the information age."

- Jake Bernstein

"Under pressure from politicians, businessmen and administrators, school districts consistently inflate scores, There are various ways to game the system."

-Jake Bernstein

placing children in special education, keeping children home on test days, and focusing teaching on kids with close-to-passing scores while ignoring those far from passing or sure to pass.

Former teacher Deborah Diffley told the *Observer*, “I’ve seen whole classes sent down the hall to watch videos while others were drilled.” Several cheating scandals have also erupted in Texas.

According to Bernstein, other test data failed to confirm TAAS gains. SAT and ACT scores of Texas high school students have been flat while other states posted increases. Only a third of all state college freshmen test-takers passed an exam intended to evaluate the skills of incoming students—most of whom had already passed TAAS.

Bernstein also quoted extensively from Texas college students regarding their disdain for TAAS (also see FairTest *Examiner*, Spring 2002). “Together the [students’] essays paint a picture of schools where ever-expanding TAAS practice forced out real curriculum and education came second to the manufacture of high test scores.”

The article particularly criticized TAAS’ impact on reading and writing: “Even the winners lose if all they get is functional literacy,” noted the *Observer*. Former teacher Julie Pennington explained, “[Y]ou give [some students who pass TAAS writing] a blank piece of paper and ask them to write a story without some kind of template, they can’t produce anything.”

“Contrary to the official line, minorities have suffered more in a TAAS-centered system,” the article reports. “‘Part of me feels like the test is in place to keep immigrant kids from succeeding,’ says a teacher who instructs mostly minority 9th- and 10th-graders in Austin on reading.”

Fear of retaliation kept Texas teachers from denouncing the system. The article used the names of retired teachers, but current teachers were anonymous. “Teachers who speak out can be charged with insubordination and fired,” the article explains.

Rather than use the Texas TAAS experiment as a template for national education reform, as NCLB has done, there are a series of

Only a third of all state college freshmen test-takers passed an exam intended to evaluate the skills of incoming students—most of whom had already passed TAAS.

Contrary to the official line, minorities have suffered more in a TAAS-centered system

cautionary lessons to be learned from Texas about relying on standardized testing to spur improvements. Among them are the following:

- Learning outcomes measured in a variety of ways do not necessarily improve even if test scores rise.
- Extreme pressure to show test score gains leads to teaching to the test, elimination of subjects that are not tested, narrowing of subjects that are taught to what is on the test, and corruption of instruction.
- Dropouts and retentions increase, particularly among low-income minority students. There is enormous pressure to hide the bad news. Rather than honest accountability, the result is Enron-style creative accounting.

B. Chicago: Behind the Model Urban District

While Texas was the state model President Bush used to promote NCLB, Chicago has often been touted as the model for urban school reform. Under a 1995 state law, Chicago School Reform Board President Gery Chico and Chief Executive Officer Paul Vallas implemented a high-stakes testing regime in which promotion from grades 3, 6 and 8 was made contingent on scores on the Iowa Tests of Basic Skills and schools could face “reconstitution” for low scores (Bryk, 2003). This was a violation of the test company’s policies regarding proper test use, which warned against reliance on exam scores to make high-stakes educational decisions (FairTest *Examiner*, 2000). Nonetheless, this test-based model became a nationally praised example of urban school reform.

Tens of thousands of children were retained in grade as test scores appeared to rise. As in Texas, appearances were deceptive. Attaching high stakes to tests is similar to holding a match to a thermostat: The numbers say the room is getting warmer, but it is not. Indeed, over time, the room will get colder while the tricked thermostat reports ever-higher readings.

Analyses of the Chicago Public Schools (CPS) by the Consortium on Chicago School Research have found that grade retention did not help and often hurt the students who were retained, including by boosting the dropout rate (Nagaoka and Roderick, 2004; Allensworth, 2004). More broadly, research by Tony Bryk (2003) of the Consortium concludes that most of the ostensible gains under centralized, test-based accountability were mere inflation - the tricked thermostat.

Attaching high stakes to tests is similar to holding a match to a thermostat: The numbers say the room is getting warmer, but it is not.

In Chicago, grade retention did not help and often hurt the students who were retained, including by boosting the dropout rate.

Bryk's article outlines a series of factors that caused test scores to rise without any real gain in student achievement. These include the following:

- Retained students were averaged into the grade they were repeating, thus boosting reported scores from the two years' classes. For example, the repeaters scored higher in grade 3, due to more familiarity with the tested materials, and grade 4 scores were higher because low scorers were held back.
- CPS changed the rules to exclude increasing numbers of students with limited English proficiency. This may have been a reasonable educational policy, but it also caused reported scores to rise.
- African-American enrollment declined while Latino enrollment grew, and those Latinos who are tested in Chicago score higher than blacks.
- More children were referred to special education, and their scores were not included.
- CPS moved the testing date to later in the year without adjusting for the change, giving scores an artificial boost.
- Chicago kept using the same three forms of the ITBS while some children were tested up to five times over 15 months. Familiarity, absent cheating or even teaching to the test, would artificially boost scores.

Bryk's study did not evaluate the impact of incessant teaching to the test or the narrowing of curriculum to tested subjects and of teaching that mirrored the tests – all of which will boost test scores without increasing learning. Worse, the narrow focus produces less learning in the long run because students are not adequately taught untested subjects nor the higher order thinking that tests cannot measure but students need to make progress in their education.

Bryk explains that virtually all the real gains on ITBS scores in the test-based accountability program occurred in 1997: "This suggests that the CPS experienced a one-time burst in student learning in the year that the high-stakes accountability was announced. However [in subsequent years]... no further productivity improvements were recognized. In fact, the annual learning gains declined in some post-1997 [years]" (p. 253). These findings are consistent with commonly found effects of narrowly teaching to the test.

The narrow focus produces less learning in the long run because students are not adequately taught untested subjects nor the higher order thinking that tests cannot measure.

If the goals of school reform are to improve overall teaching and learning and close academic achievement gaps, test-based accountability has been a failure in Chicago.

Bryk also found that African-American and Latino students fared no better than whites under test-based accountability. If the goals of school reform are to improve overall teaching and learning and close academic achievement gaps, test-based accountability has been a failure in Chicago.

Fortunately for Chicago students, the new leadership of CPS has somewhat de-emphasized the use of tests for grade promotion and has begun to chart a new course toward school improvement that focuses more on teacher professional development, including the use of classroom-based assessment. It is far too early to determine how well these initiatives are being carried out, their ultimate success, or whether CPS leadership will find ways to prevent the damage caused by high-stakes testing for schools as now mandated by NCLB.

C. Why NCLB Will Hurt Rather Than Help Teaching and Learning

A major rationale for NCLB's mandated increases in standardized testing is that learning problems must be identified by testing before they can be solved. Once identified, the theory goes, teachers can better focus on problem areas and improve students' understanding and achievement.

Unfortunately, there is little evidence to support the theory that a radical increase in standardized testing and intensified stakes will improve the quality of teaching and learning. Looking beyond the cautionary tales from Texas and Chicago, a mounting body of evidence shows that when the gauge of student progress is reduced to reading and math test scores, schools tend to narrow instruction to what is tested. Education is damaged, especially in low-income and minority schools, as students are coached to pass narrow tests rather than learning a rich curriculum to prepare them for life in the 21st century.

Proponents have tried to distinguish NCLB from the burgeoning controversy surrounding high-stakes testing because it is "not high stakes for students." The fact remains that NCLB does impose extremely high stakes on schools and teachers. The pressures and stresses entailed are likely to be passed along to students, particularly disadvantaged groups of students, in many ways. If anything, by putting pressure directly on schools and especially on teachers in tested subjects, narrow teaching to the test will continue to increase.

There is little evidence to support the theory that a radical increase in standardized testing and intensified stakes will improve the quality of teaching and learning.

By putting pressure directly on schools and especially on teachers in tested subjects, narrow teaching to the test will continue to increase.

According to a recent study by the Center on Education Policy, 52 percent of public school students nationwide and a greater share, 55 percent, of minority public school students live in states that now have exit exams (Gayler *et al.*, 2003). Some of these states will use their high-stakes exams to comply with NCLB mandates. Though the growth in the number of states with high stakes for individual students has stalled, adoption of such exit exams could be accelerated by the rationale that more pressure must be applied to students if they are to take the tests seriously enough for the school to avoid being labeled INOI and sanctioned (Stecher *et al.*, 2003). Kentucky is one state that is considering imposing stiffer penalties on students who perform poorly to counter disappointing progress on state tests (AP, 2003).

The Failure of High Stakes

Whether punitive consequences are imposed directly on students or indirectly on their schools, NCLB has many of the hallmarks and likely negative consequences of other high-stakes testing systems, including:

- Despite widespread adoption of “school reforms” based on high-stakes testing, there is little evidence that the model works. A series of studies have looked at academic achievement in states with high-stakes systems and found little evidence of increased learning, and in some cases a decline. Whatever small gains might exist are nowhere near the rate of progress required by NCLB.
- Teacher surveys and other research demonstrate that an overemphasis on test results for accountability promotes excessive teaching to the test and dilution of the curriculum. There is also evidence that this narrowing and dumbing down of the curriculum is most intense in lower-performing schools that serve low-income and minority students, raising critical questions of equity.
- Test-based reform advocates claim that as long as state tests are aligned with high-quality standards, teaching to the test ensures that students are being taught a high-quality curriculum. However, independent analysts have found that tests, such as the New York State Regents Exams, often fail to measure the standards and objectives deemed most important. Teaching to tests that are devoid of the most important standards implies that students will not be exposed to high-quality curricula.

Whether punitive consequences are imposed directly on students or indirectly on their schools, NCLB has many of the hallmarks and likely negative consequences of other high-stakes testing systems.

Teaching to tests that are devoid of the most important standards implies that students will not be exposed to high-quality curricula.

- When tests control classrooms, the quality of teaching tends to suffer because it is assumed that all students who fail need the same kind of remedial instruction. On the contrary, researchers have found students fail for different reasons and need different instructional approaches to get on track.
- An essential premise of high-stakes testing is that the stakes are a key motivating factor, causing students to study harder and learn to higher levels. Research done in the U.S. and Great Britain, however, found little evidence to support the premise. A particularly troubling finding is that low-achieving students are most likely to become discouraged in a high-stakes testing environment.
- A growing body of data shows that test-based reform policies are linked to falling graduation rates as well as to evidence that states conceal how many students drop out or are pushed out if they struggle to achieve on state tests. There is also evidence that schools are retaining more students in hopes of reaping higher test scores in key grades, while putting the retained students at a higher risk of dropping out of school.
- When applied to English language learners, NCLB's mandates create a catch-22 whereby students are no longer counted as limited English proficient once they meet the test's standards, so schools can never claim credit for their improvement
- Many educators and parents view the demand that disabled students match the learning pace and level of achievement of their nondisabled peers on state tests as an impossible goal that will cause increased scapegoating.
- While early childhood educators caution against using standardized testing with young students, there appears to be a trend toward testing children as young as kindergarten and even preschool, in preparation for NCLB accountability, with possibly damaging consequences.

Does the Evidence Support the Model's Effectiveness?

Proponents defend the NCLB approach by claiming that testing is just one component of a model that includes standards and either rewards or punishment for achievement. The question is whether there is any convincing evidence that this model, which in various forms has been used across the U.S. for years, works to improve educational outcomes.

Low-achieving students are most likely to become discouraged in a high-stakes testing environment.

When applied to English language learners, NCLB's mandates create a catch-22 whereby students are no longer counted as limited English proficient once they meet the test's standards.

A recent report from the RAND Corp. (Stecher *et al.*, 2003), looked at a range of research and found the evidence of the model's success limited and inconclusive:

“Although there is an appealing logic to the idea that high standards, testing, and consequences have significant power to shape educators' behavior and to improve student learning, there is only limited evidence about their effectiveness. On the plus side, schools, teachers, and students seem to respond to the incentives created by accountability systems, and scores on state tests typically rise after the system is introduced. There is also evidence that scores on some external tests, such as the NAEP, may rise when states implement accountability systems (Carnoy and Loeb, 2002). On the minus side, higher test scores do not necessarily reflect real gains in student mastery of content standards; rather, for example, they may reflect students' learning of particular test content or formats. Even when NAEP scores rise, the gains on NAEP tend to be many times smaller than the gains on the state test of the same subject matter (Linn, 2000; Koretz and Barron, 1998). One interpretation of this difference is that the high stakes associated with the state test led to inflated scores, while the smaller gains on the low-stakes NAEP are more indicative of students' true performance.”

Even if NAEP scores show a modest increase in some states with high-stakes testing, this may be a consequence of intense focus on reading and math, to the detriment of other subjects. While proponents justify this narrow focus as a return to the basics, upon which a broader education can be built, others—including most parents, according to polling data—expect their children to receive a broader and more balanced curriculum and worry that testing in English and math will narrow what their children learn (Rose and Gallup, 2003).

The authors of a recent study of high-stakes accountability systems in Southern states point out that while proponents tend to link any student gains to standards-based reform, there may be a host of other factors that are equally important or more so. In North Carolina, for example, researchers cite “a long-term investment in the quality of the teaching force and support for teachers' work [that] began well before standards, assessment, and accountability structures were fully built, and this investment appears central to the state's successes to date” (Berry *et al.*, 2003).

Even if NAEP scores show a modest increase in some states with high-stakes testing, this may be a consequence of intense focus on reading and math, to the detriment of other subjects.

While proponents tend to link any student gains to standards-based reform, there may be a host of other factors that are equally important or more so.

Studies that set out to gauge the effects on teaching and learning support the contention that high-stakes testing has had a more negative than positive impact overall. Calling high-stakes testing “a failed policy” with discriminatory impacts, Arizona State University (ASU) researchers Audrey Amrein and David Berliner (2002) conclude, “While a state’s high-stakes test may show increased scores, there is little support in these data that such increases are anything but the result of test preparation and/or the exclusion of students from the testing process.”

The authors examined 18 states that have implemented graduation exams and other high-stakes testing of students. The data used in the analysis consisted of scores obtained over two decades from four commonly used standardized tests: the ACT, SAT, National Assessment of Educational Progress (NAEP) and Advanced Placement (AP) tests. The study investigated whether students demonstrated any transfer of knowledge beyond what was needed to perform on the state’s own high-stakes test. If scores on other tests increased following implementation of a high-stakes testing program, it would be evidence that this approach promotes transfer of learning.

The study found the policy “is not working,” concluding, “In all but one analysis...student learning is indeterminate, remains at the same level it was before the policy was implemented, or actually goes down when high-stakes testing policies are instituted.” For example, average scores on the ACT, the college admissions exam most commonly used in states with high-school graduation exams, declined in 67 percent of the states requiring exit tests. In addition, the proportion of students taking the ACT grew more slowly in the high-stakes states than in the nation as a whole, suggesting that the graduation tests are not encouraging more students to attend college. The results are similar for the SAT: 56 percent of the states with graduation tests experienced average SAT score declines, and SAT participation rates fell in 61 percent of the states, when compared with the nation as a whole.

Overall, NAEP math and reading results at grades 4 and 8 had no correlation with the existence of high-stakes tests. (NAEP does not report state-level scores for grade 12). In reading, students in high-stakes states did improve slightly more from grade 4 in 1994 to grade 8 in 1998 than did the nation as a whole. But this was the only finding in the study to lend any support to proponents of high-stakes tests.

Audrey Amrein and David Berliner (2002) conclude, “While a state’s high-stakes test may show increased scores, there is little support in these data that such increases are anything but the result of test preparation and/or the exclusion of students from the testing process.”

The study found the policy “is not working,” concluding, “In all but one analysis...student learning is indeterminate, remains at the same level it was before the policy was implemented, or actually goes down when high-stakes testing policies are instituted.”

The researchers also found that states with high-stakes exams are more likely to exclude students with disabilities or limited English proficiency from participation in NAEP. This largely explained the greater-than-average score gains in the high-stakes states of Texas and North Carolina.

When rates of participation in the AP program were controlled, there was a decrease in the percentage of students scoring “3” or higher on AP tests in most states with high-stakes graduation exams. Participation rates in AP fell in 67 percent of the high-stakes states, compared with the national average.

In sum, in states with high-stakes graduation tests, scores on independent exams provide no evidence of increased student learning and often decline relative to the nation as a whole. Many students appear less well prepared and less likely to go to college than their peers in non-high-stakes states. Amrein and Berliner suggest this may be because high-stakes testing leads to narrow training to help students pass specific exams, not to education that leads to genuine learning.

In part because states with high-stakes graduation tests are poorer and have larger proportions of minority-group students, the researchers point out that the damage more often affects these students than their wealthier, majority-group peers. Thus, they conclude, “a high-stakes testing policy is more than a benign error in political judgment. It is an error in policy that results in structural and institutional mechanisms that discriminate against all of America’s poor and many of America’s minority students.”

Teaching to the Test

An extensive survey of teachers about the impact of state tests on their classrooms bolstered the claim that high-stakes testing causes instruction to be narrowed to prepare students for specific state tests. The National Board on Testing and Public Policy at Boston College (Pedulla *et al.*, 2003; Clarke *et al.*, 2002) reported that three-quarters of surveyed teachers said state testing programs were not worth the time and money. A substantial majority said testing caused them to teach in ways that contradicted their views of sound instruction.

The board released two studies of teachers’ views of the effects of state-mandated testing on teaching and learning, one a national sur-

In sum, in states with high-stakes graduation tests, scores on independent exams provide no evidence of increased student learning and often decline relative to the nation as a whole.

Three-quarters of surveyed teachers said state testing programs were not worth the time and money.

vey, the other from in-depth interviews. In both studies, teachers said higher stakes created more pressure to teach to the test. About 40 percent of survey respondents said students could raise their test scores without improving their real knowledge. As stakes increased, teachers were more likely to narrow classroom curriculum to focus on tested areas and to engage in more test preparation, including use of items similar to those that are on the exams.

The interviews were conducted in Kansas (low stakes), Michigan (medium) and Massachusetts (high). As stakes increased, so did teachers' reports of test-related effects on their classrooms. Some findings:

- Only one in ten urban Michigan teachers thought the state's test-based scholarship awards motivated their students, while just one-third of suburban and rural teachers did.
- In Massachusetts, more than half the high school teachers thought that testing demoralized their students. Two-thirds of all teachers thought the tests were unduly stressful and unfair to special populations. Four out of five thought the exam should not be used as a sole hurdle for graduation.

"Teaching to the Test: The Good, the Bad, and Who's Responsible," by Nancy Kober (2002), looked into the widespread concern that high-stakes testing causes "teaching to the test." Kober concluded that teaching to the test can cover a range of activities, some clearly harmful, such as outright cheating. A positive example of teaching to the test, which she notes is rare, would be to focus instruction on the most important knowledge and skills contained in high-quality state standards and measured by well-designed tests.

Kober explained that while teachers are often blamed, state and national policymakers should really be held accountable for inappropriate forms of teaching to the test. "These leaders have created accountability systems centered on higher test scores, with little regard for how these scores are attained." NCLB's testing provisions, she said, will increase pressure on teachers and students even more.

A study released in March 2004 appeared to support the prediction that NCLB would create more pressure to narrow curricula to what is tested and eliminate important subjects. The Council for Basic Education studied the effects of NCLB on instructional time and revealed that subjects like social studies, civics, geography, languages and the arts are being given short shrift because of increasing time

In Massachusetts, more than half the high school teachers thought that testing demoralized their students. Two-thirds of all teachers thought the tests were unduly stressful and unfair to special populations.

*"These leaders have created accountability systems centered on higher test scores, with little regard for how these scores are attained."
-Nancy Kober*

devoted to reading, math and science (von Zastrow, 2004). “The narrowing of the curriculum is worrisome because students need exposure to history, social studies, geography, and foreign languages to be fully prepared for citizenship, work, and learning in a rapidly changing world,” said Raymond “Buzz” Bartlett, president of the Council for Basic Education. “Truly high expectations cannot begin and end with math, science, and reading.”

A group of school administrators reacted to the study by saying it confirmed their concerns about NCLB. “Sadly, the survey findings reflect our fear that schools are being forced by NCLB to ‘teach to the test,’” stated American Federation of School Administrators National President Baxter Atkinson. “When you teach to the test, you place a greater emphasis on the subjects being tested, and a lesser emphasis -- including the elimination of -- the subjects not being tested. Students need a well-rounded curriculum. NCLB is having a detrimental effect on curriculum, and ultimately on students.”

News reports describing such narrowing are beginning to accumulate. In one Indiana middle school (Dobbs, 2004), the drive to raise math and English scores has meant the elimination of two arts teachers, home economics, most foreign-language classes and some physical education classes. The theory is that without these “extras,” students can spend four or five hours per day on reading and math. In practice, the *Washington Post* reported, “it is hard to retain their attention for more than a few minutes. On a recent day, one student was playing video games on a computer at the back of the classroom while [the sixth-grade teacher] was threatening to send another to the principal for disruptive behavior.”

The Council for Basic Education study also bolstered the contention that minority students are most likely to have their schooling narrowed by the pressures of high stakes. The authors found that nearly half (47 percent) of principals at schools with high minority student enrollments reported decreases in elementary social studies; four in ten (42 percent) anticipated decreases in instructional time for the arts; and three in ten (29 percent) foresaw decreases in instructional time for foreign language.

A study focusing on high-stakes accountability systems in 24 schools in Southern states looked at the link between high-stakes accountability and professional development. The report by Barnett

“Students need a well-rounded curriculum. NCLB is having a detrimental effect on curriculum, and ultimately on students.”

-Baxter Atkinson

The Council for Basic Education study also bolstered the contention that minority students are most likely to have their schooling narrowed by the pressures of high stakes.

Berry and his colleagues (2003) characterized as “evident and profound” the different consequences for higher versus lower performing districts and schools. The authors concluded that high-stakes accountability systems were more likely to make teachers in low-performing districts feel pressure to narrow the curriculum. “We also found that the pressure has also resulted from teachers who are overworked and exhausted from trying to cover increasing amounts of content and then having to teach to the test.”

The authors concluded that narrowing of curricula is “an almost inevitable” result of high-stakes accountability, with science and social studies dismissed while math and English assume centrality. Here is what the results looked like to one Tennessee teacher:

“[She said] she was expected to teach more and more facts as the state expands the curriculum, but that no one ever removes any items from the curricular frameworks. She reported feeling like she was on a ‘treadmill, pushing students harder and harder.’ Her school and others have begun to eliminate recess (even for young elementary age children), cut back on art and music, and cancel field trips all [sic] in the name of spending more time on preparing students for the tests” (Berry, *et al.*, 2003).

When RAND researchers Stephen P. Klein and his colleagues (2000) looked at Texas’s experience with the TAAS, they also found evidence that high-stakes accountability systems narrow the curriculum. “States that use high-stakes exams may encounter a plethora of problems that would undermine the interpretation of the scores obtained,” the authors said, including, “narrowing the curriculum to improve scores on the state exam at the expense of other important skills and subjects that are not tested.”

A more recent Rand report questioned some of the key assumptions of NCLB, including the idea that incentives will not encourage corruption of the system. “Researchers have found quite the opposite, i.e., that accountability systems that use a limited number of quantitative indicators lead staff to focus narrowly on measured outcomes at the expense of other goals” (Stecher, *et al.*, 2003).

For some teachers, this stress leads them to flee the schools and children who arguably need them the most, as did one teacher who left a low-performing school because of pressure to teach to the test. “It

Pressure has also resulted from teachers who are overworked and exhausted from trying to cover increasing amounts of content and then having to teach to the test.

The authors concluded that narrowing of curricula is “an almost inevitable” result of high-stakes accountability, with science and social studies dismissed while math and English assume centrality.

was all drill, drill, drill for the test,” said Lana Curtis, now a sixth grade teacher in prosperous Roanoke Rapids, NC. “I did not feel that I was being treated as a professional. Pretty much everything we taught was related to the test” (Dobbs, 2003).

Writing in the *Washington Post*, Emmet Rosenfeld (2004) described how he was driven from public school by the pressure to tailor his teaching to fit the multiple-choice Virginia Standards of Learning tests: “[I]t pained me deeply to find myself in a situation where I felt compelled to give a rarely engaged student a practice bubble test instead of letting him read a book he had discovered he loved.”

Berry, *et al.* (2003) also suggest that the same lack of resources, leadership and capacity that make schools low performing in the first place make it unlikely that they will use data from accountability systems to improve student performance. Their work underscores the folly of imposing expensive accountability systems on schools that lack sufficient resources for improvement. Without this capacity, the low-performing schools are under the gun to do whatever they can to prepare students for the tests.

John Diamond and James Spillane of Northwestern University (2002) also found that the response to high stakes accountability in low-performing schools may be counterproductive. They closely examined four Chicago schools—two performing fairly well on mandated tests, two doing poorly. All the schools did considerable test preparation, but the lower-performing schools tended to test more, focus attention on those close to passing, and engage in other activities not likely to help most of their students. Since low-income and minority-group students are concentrated in lower-performing schools, which rely on unhelpful methods while schools serving wealthier students use more effective methods to raise scores, high-stakes accountability testing could widen not only test-score gaps but also widen the gaps in real learning opportunities.

Another recent study showing classroom damage is Tonya Moon *et al.* (2003), who conclude, “students from poverty are less likely to be exposed to challenging curricula and instructional methods... [so that] accountability through testing is a vehicle to restrict educational opportunities from those who need opportunities most.”

“[I]t pained me deeply to find myself in a situation where I felt compelled to give a rarely engaged student a practice bubble test instead of letting him read a book he had discovered he loved.”

-Emmet R.
Rosenfeld

Students from poverty are less likely to be exposed to challenging curricula and instructional methods... [so that] accountability through testing is a vehicle to restrict educational opportunities from those who need opportunities most.

What Do the Tests Really Measure? The Problem of ‘Alignment’

The test-based reform and accountability model presumes that tests will be aligned with high-quality standards and objectives. If this were true, proponents could rightly claim that when teachers teach to the test, they are teaching to “high standards.”

A recent RAND study emphasized the importance of alignment of standards and tests. “For test scores to provide a meaningful indication of proficiency according to state standards, the tests must reflect the breadth and depth of the standards. However, satisfying this criterion is far more difficult than it appears. While it is relatively easy to make a superficial match between test questions and standards, it is extremely difficult to determine whether the skills needed to answer a question or a group of questions are the same as the skills described in a standard” (Stecher *et al.*, 2003).

An analysis of state tests by Robert Rothman and his colleagues for the National Center for Research on Evaluation, Standards and Student Testing (CRESST) also found troubling patterns in the way tests are aligned with standards. In “Benchmarking and Alignment of Standards and Testing,” Rothman *et al.* (2002) found it positive that states are trying to align tests with their standards, items are limited to material in the standards, and individual items are generally well aligned to the standard to which they are mapped.

However, they found *that test items do not assess the standards and objectives deemed most important by the state*. Challenging content is underrepresented or omitted; questions tend to measure simpler cognitive processes such as routine math calculations. “This particular pattern of non-alignment can have serious consequences for the kind of teaching that will occur in the states using such tests,” the report says. “As test-based accountability becomes more stringent, schools and teachers will match their curriculum and teaching ever more closely to what is on the tests, rather than to what the standards say ought to count. The result will be an increasing focus on the low-demand aspects of the state’s standards and a decreasing focus on the high-demand aspects that define a rigorous curriculum.”

Even Achieve, Inc. (2001), an organization dedicated to promoting standards and tests, has concluded that state exams generally fail to adequately assess state standards. A discussion of Massachusetts state

While it is relatively easy to make a superficial match between test questions and standards, it is extremely difficult to determine whether the skills needed to answer a question or a group of questions are the same as the skills described in a standard”

-Stecher *et al.*,

Test items do not assess the standards and objectives deemed most important by the state.

tests acknowledges that standardized tests are poor tools for measuring many of the more complex skills and higher order thinking expected in the standards (Achieve, 2001). Similarly, researchers at the University of Wisconsin found a generally poor match between state standards and tests, with tests focusing on lower-level, easier-to-measure knowledge and skills (WCER, 1999).

If the goal is to prepare and create more opportunities for students to do well in college, the implications of this failure to assess higher order thinking are troubling. A recent study by University of Oregon's Center for Educational Policy Research (Conley, 2003) adds fuel to such concerns. The study looked at 35 English/language arts exams and 31 mathematics tests from 20 states and found that often they did not gauge students' readiness for college-level work.

Tests Worth Teaching To? The Case of the New York Regents

The New York State Regents exams, wracked by controversy over high failure rates, are a prime example of widely praised and highly rated tests that failed to hold up under scrutiny. Starting in the fall of 2001, five panels of writers, journalists, academics, and college admissions officers have met under the auspices of the Rockefeller and Soros Foundations, working as the Center for Inquiry (2001) to analyze the mandatory Regents graduation exams. The panelists took and then discussed the tests, which are made public after administration. The panels sharply criticized the quality of the tests and pointed out that teaching to them will inhibit high-quality learning. The following are comments from the first four reports:

English Language Arts:

- The short-answer, multiple-choice questions... were insulting to the literature, the author, and the student.
- It seems that the framers of this exam have somehow managed to be ignorant of every piece of research that has been published in the last 25 years about rhetoric and the writing process. While taking this exam, we had to forget about everything we know or have learned about writing.

These panelists, all published authors, advised test-takers not to think, but instead to learn test-taking tricks.

The New York State Regents exams, wracked by controversy over high failure rates, are a prime example of widely praised and highly rated tests that failed to hold up under scrutiny.

These panelists, all published authors, advised test-takers not to think, but instead to learn test-taking tricks.

Global History and Geography:

- Students could be well prepared for the test not by studying content but by learning “test tricks.”
- One of the journalists observed that this exam... required students to be uncritical.
- The essay question did not allow a student to demonstrate an ability to evaluate contradictory sources, since all the sources suggested a single possible answer to the question.

English Language Arts (ELA) and Global History (reviewed by college admissions officers):

- Our panel concluded, “If you want to know whether this test helps prepare kids for college, the answer is no.” College requires critical thinking and the weighing of evidence; this test does not. As one of our participants noted, these tests “simply test how well people can take tests.”
- There is “very little predictive validity between doing well on this test and doing well in college.”

Living Environment (Science):

- A number of panelists who work with high-school students daily emphasized that the exam’s focus on memorizing detailed content will not engage students... (who) will perceive the content... as “abstract, useless, and boring.”
- We were largely in agreement that a good test-taker could nevertheless have an astonishingly poor understanding of science.

High-Stakes Testing Weakens Instruction

Contrary to the premise that test results will identify students and areas that they need help in so that teachers can address weaknesses, there is evidence that the quality and efficacy of curriculum and instruction tend to suffer when tests control curriculum and instruction.

Two scholars looked at instructional responses to high-stakes test results (Riddle Buly and Valencia, 2002), using Washington state’s fourth-grade reading test. They concluded that students who struggle the most often receive instruction that is not geared to their individual needs and unlikely to help them.

Sheila Valencia, a professor in the University of Washington’s College of Education, and Marsha Riddle Buly, an assistant professor at

“If you want to know whether this test helps prepare kids for college, the answer is no.”

“College requires critical thinking and the weighing of evidence; this test does not.”

-College Admissions Officers

There is evidence that the quality and efficacy of curriculum and instruction tends to suffer when tests control curriculum and instruction.

Western Washington University, found that students failed the reading test for a wide variety of reasons. Too often, however, school districts responded to such failures with standardized instructional approaches.

“People are trying to put into place instructional programs that they think will help low-achieving kids,” Valencia said. “They’re making the assumption that all kids who failed to meet a standard need the same kind of instruction — if you fail, you should get this kind of program. We think that’s problematic.”

Valencia and Riddle Buly said their data points to the need for “policies that require multiple indicators of achievement, a stance that has long been advocated but rarely implemented.” To avoid the mistake of standardized responses to nonstandard learning issues, there should be “complex indicators of student performance in the targeted subject area,” which they said is information that can only come from classroom-based assessments, not standardized state exams.

A number of other studies have similar findings, including Moon *et al.* (2003), Hinde (2003), and Mabry *et al.* (2003).

Tests Demotivate

A key claim of high-stakes testing proponents is that such tests motivate students to work harder and thus learn more. But again, there is substantial data that suggests the opposite relationship may be true, particularly for students who have not been succeeding in school.

A thorough summary of research on education and motivation by a British team found that constant testing motivates only some students and increases the achievement gap between higher and lower achieving students. The study, titled “A Systematic Review of the Impact of Summative Assessment and Tests on Students’ Motivation for Learning,” rebuts the claim that standardized testing motivates low achievers to reap the reward of high scores and avoid the punishment of failure. In fact, researchers Wynne Harlen and Dr. Ruth Deakin-Crick (2002) of Bristol University found that the two categories of students particularly discouraged by constant testing are girls and low achievers.

These findings call into question the claims of U.S. high-stakes testing proponents that they have found the key to closing the race-

Valencia and Riddle Buly said their data points to the need for “policies that require multiple indicators of achievement, a stance that has long been advocated but rarely implemented.

Constant testing motivates only some students and increases the achievement gap between higher and lower achieving students.

based achievement gap. The results suggest that groups such as low-income and minority students, who traditionally score low on standardized tests, are likely to be among those who are discouraged by consistently poor test results.

The study also found that constant testing encourages even successful students to see the goals of education in terms of passing tests rather than developing an understanding of what they are learning. The researchers found firm evidence that achievement of literacy is linked to students' interest in learning, the degree to which their learning strategies link to existing knowledge rather than just memorizing, and the degree to which they feel in control of their learning. The latter point confirms a rarely noted finding of the famous Coleman report (1966) that the only factor significantly influencing learning outcomes, in addition to family and social background, is students' sense of control over their learning.

The British study echoes the findings of a U.S. review of the research by Richard Ryan (1999). Ryan's comprehensive analysis of research on motivation concluded that test-based reform damages rather than improves the processes through which higher quality academic achievement occurs. This has occurred because reformers have defined achievement too narrowly as higher test scores. The "hidden costs" of this approach, Ryan says, are "lowered interest for school in both teachers and students, lower quality learning, and the development of motivational orientations that, again paradoxically, will not aid students in their future endeavors." Teachers must have the freedom and resources to focus on individual learners and their unique abilities if the aim is to avoid leaving behind students who have alternative learning styles and interests and to obtain a higher quality of achievement for all.

Author Alfie Kohn (1999) has also exhaustively examined research on motivation. He has reported that high-stakes tests reduce intrinsic motivation and turn eagerness to learn into an eagerness to beat the test for the sake of a reward. In the long run, the evidence shows, the emphasis on testing produces people with less real interest in learning.

Constant testing encourages even successful students to see the goals of education in terms of passing tests rather than developing an understanding of what they are learning.

In the long run, the evidence shows, the emphasis on testing produces people with less real interest in learning.

D. Dropouts: NCLB Seen Likely to Add to the 'Disconnected'

A recent study by Northeastern University Center for Labor Market Studies economist Dr. Neeta P. Fogg found large numbers of young people between the ages of 16 and 24 are “disconnected”: They’re not in school, they don’t have jobs, and their future is bleak. In Chicago, Dr. Fogg found, 22 percent are not connected to school or work. Most shocking is the disparity between whites and blacks. “In the city of Chicago,” Fogg wrote, “Black males were nearly 6 times more likely to be out of school and out of work compared to White males. Nearly 45 percent of Black male residents of the city between the ages of 20 and 24 were out of school and out of work compared to less than 8 percent of White male residents of the city.” Fogg noted that this is typically the time when young people “accumulate human capital in the form of formal education attainment or work experience in the labor market” (Fogg, 2003).

Fogg drew no link between her findings and the effects of testing policies like NCLB. But when he reported on the study in the *New York Times* (2003), columnist Bob Herbert concluded that NCLB is likely to worsen the situation. “The recent increased federal involvement in the nation’s public schools is having the perverse effect of driving up dropout rates as school administrators try to pump up their high-stakes test results by getting rid of struggling students.”

There is substantial data to bolster Herbert’s conclusion, and evidence as well that states are increasingly at pains to conceal the extent of the problem. A recent report from the conservative Manhattan Institute by author Jay Greene, for example, found three in 10 high school freshmen and half of all black and Latino students never make it to graduation, even though many states report dropout rates in low single digits. “We’re amazed at official statistics that look very far off from what the truth must be. We are seeing fudging in a lot of places,” Greene said. The gap was largest in South Carolina and Texas, both of which he said undercounted dropouts by roughly 30 percent.

At a New York legislative hearing on problems with the state’s Regents exams, Boston College Professor Walt Haney testified: “The rate at which students are graduating from New York high schools has been plummeting and the Empire State now has one of the worst graduation rates of any state in the nation. Rates at which Black and Hispanic students are graduating from high school in New York are

“In the city of Chicago, Black males were nearly 6 times more likely to be out of school and out of work compared to White males.”

-Neeta P. Fogg

“The recent increased federal involvement in the nation’s public schools is having the perverse effect of driving up dropout rates as school administrators try to pump up their high-stakes test results by getting rid of struggling students.”

- Bob Herbert

shockingly low. Rates at which students in New York are missing between grade 9 and 10 have been increasing sharply over the last decade, a condition that bodes ill for their likelihood of persisting in school to graduation” (Haney, 2003).

Haney and his colleagues subsequently released a report that looked at graduation rates nationwide and found that they fell in the 1990s as states instituted accountability systems similar to those in NCLB. The report, “The Education Pipeline in the United States, 1970-2000,” compares school enrollment data by grade from the Education Department’s National Center for Education Statistics. It found that on-time graduations declined by four percentage points, to 74.4 percent in 2000-01 from 78.4 percent in 1991-92. Haney links the decline to increasing course requirements and increasing demands that students pass state exams to receive a high school diploma (2004).

The study found that over the last 30 years, the rate at which students disappear between grades 9 and 10 has tripled. Between the 1999-2000 and 2000-2001 academic years, nine of the 12 states with the worst declines in numbers of students moving from grade 9 to grade 10 were in the South, where high-stakes testing policies have long been embraced. The other states were Nevada and New York, which have high-stakes graduation tests, and Hawaii.

“These findings,” said Haney, “are quite disturbing. Despite all of the high-sounding rhetoric about reforming our schools, the data on enrollment and graduation demonstrate that many states hold students back in 9th grade, encourage dropping out, and graduate a declining percentage of students.”

Some Drop Out, Others are Pushed

In “Pushing Out At-Risk Students: An Analysis of High School Discharge Figures,” the New York group Advocates for Children (2002) reported on the growing push-out phenomenon in New York public schools. “An alarming number of high school students are leaving New York City public schools without graduating,” the report said. “It appears that in some cases school officials are encouraging students to leave regular high school programs even though they are of school age or have a right to receive appropriate literacy, support, and educational services through the public school system.”

Over the last 30 years, the rate at which students disappear between grades 9 and 10 has tripled.

“Despite all of the high-sounding rhetoric about reforming our schools, the data on enrollment and graduation demonstrate that many states hold students back in 9th grade, encourage dropping out, and graduate a declining percentage of students.”

-Walt Haney

According to the report, prepared in conjunction with the Office of Public Advocate Betsy Gotbaum, more than 160,000 high school age students were discharged from New York City public schools during the 2000-2001, 1999-1998, and 1998-1997 school years.

The New York Times also gave prominent coverage to the push-out phenomenon in a two-part series, detailing how students who are struggling academically are pushed out of school and classified under bureaucratic categories that hide their failure to graduate (Lewin and Medina, 2003).

There are reports of pushouts in other cities. For example, the *Orlando Sentinel* reported in March that 126 students were dropped from the roll of Oak Ridge High School in Orlando last year in the weeks before the state test, the FCAT. The paper reported that after losing these low-performing students, the school's state accountability rating – based on test scores — rose from an F to a D rating. It was not clear whether these students returned to school after the test period, but if they were cut from school without their parents' permission, it would have been a violation of state law (Shanklin, 2004).

NCLB does require states to report and show improvement in graduation rates. However, states are not required to reach 100 percent graduation or meet any mandated rate of improvement.

Federal officials have approved a wide range of methods for states to determine graduation rates. As a result of this “flexibility,” there is no way to compare states' progress or lack thereof toward the goal of higher graduation rates.

The Urban Institute looked at the issue of graduation rates and NCLB. Their study used three methods of calculating graduation rates, all of which comply with federal guidelines, and found that, with one method, 20 out of 24 states would meet a performance standard of a 75 percent graduation rate. With the other two methods, only eight or nine states would meet the goal (Swanson, 2003).

Unless dropout analysis is done with sufficient rigor, NCLB could reward behavior like that of Houston. In other words, the appearance of increased graduation rates could pass muster even if the reality is that dropout rates are on the rise. It is too soon to say whether or not

The Orlando Sentinel reported in March that 126 students were dropped from the roll of Oak Ridge High School in Orlando last year in the weeks before the state test, the FCAT.

Unless dropout analysis is done with sufficient rigor, NCLB could reward behavior like that of Houston. In other words, the appearance of increased graduation rates could pass muster even if the reality is that dropout rates are on the rise.

this graduation rate provision will be a countervailing force against pressures to push out low scorers, but there is evidence that school systems have and will continue to find their way around such requirements.

Evidence of Links between High Stakes and Increased Dropouts

In the second part of their national study, Arizona State University researchers Audrey Amrein and David Berliner (2002) examined dropout and graduation rates in the 16 states that used high school exit exams in the 1990s. They found that the graduation rate decreased in 10 states after high school exit exams were implemented and increased in only five states. Similarly, dropout rates increased in eight states and decreased in five. They also found that General Equivalency Diploma (GED) enrollments tended to increase and the age of GED examinees decreased in states with exit exams, indicating that more students had left school before graduating.

To consider other consequences, the authors also examined news clips from 26 states with high-stakes tests for students or schools. The authors found tendencies toward greater grade retention (a policy that fails to improve student learning while harming children); more student expulsions, in some cases apparently to drive out low scorers; and increased exemptions of students with disabilities or limited English proficiency.

Researcher Brian Jacob (2001) also conducted an extensive national study on the question of whether graduation tests produce an increase in dropouts. In “Getting Tough? The Impact of High School Graduation Exams,” Jacob found that students who scored in the bottom fifth on tests in the eighth grade were 25 percent more likely to drop out in states that had graduation exams. The tests, however, “have no appreciable effect on the probability of dropping out for the average student.” Jacob also found that minimum competency tests (MCTs) had no overall impact on student achievement, but students in the bottom ten percent in grade 8 had smaller reading gains in states or schools with graduation tests. Jacob used data from the National Educational Longitudinal Survey (NELS). He studied the impact of MCTs from 1988 to 1992, not the more difficult graduation tests some states have recently introduced.

The graduation rate decreased in 10 states after high school exit exams were implemented and increased in only five states.

Students who scored in the bottom fifth on tests in the eighth grade were 25 percent more likely to drop out in states that had graduation exams.

Retention Could Rise

Another means by which high-stakes testing increases the dropout rate is through increased grade retention. Often touted as good for students because it ends “social promotion,” in fact retention has been exhaustively documented as having almost entirely negative consequences (Shepard & Smith, 1989; Heubert & Hauser, 1998). Nevertheless, a few states have begun to mandate that students who fail a test be retained. This policy is also being pursued in some large cities, even though previous uses of test-based retention in large cities were dropped as disasters (Moore, 2003).

More commonly, school systems respond to the pressure to raise scores by retaining students. This prevents low scorers from reducing test averages in the next higher grade. In addition, the retained students are likely to score higher when they repeat the test. This makes schools look better and appears to show that retention works.

However, retained students fall behind once they are advanced, so all the supposed benefits of test score gains are soon lost. This makes them candidates for being retained again and repeating the cycle. Not only do retained students not show sustained learning gains, but students who are retained once are twice as likely to drop out, and students who are retained two or more times become quite unlikely to finish school at all (Shepard & Smith, 1989; Heubert & Hauser, 1998).

NCLB, with its overwhelming pressure to boost scores, makes this failed policy all the more likely to be implemented because of its short-term, score-boosting impact. This winter, New York City Mayor Bloomberg and schools Chancellor Klein proposed a test-based grade three retention policy. The proposal aroused a firestorm of opposition. The City had such a policy in the 1980s, Promotional Gates, which failed to improve test scores but drove up the dropout rate. Under NCLB, totally failed policies are recycled by those who fail to learn from the past.

Often touted as good for students because it ends “social promotion,” in fact retention has been exhaustively documented as having almost entirely negative consequences.

E. The Limits of NCLB for Limited English Proficient Students

Limited English proficient (LEP) students are identified as one of the subgroups on which NCLB is focused. It is undeniable that LEP students have historically been neglected in the U.S. education system and deserve the attention of policymakers. Yet the reliance on standardized testing and punishments is particularly disastrous for this group of students.

Failure of LEP students to exhibit adequate yearly progress in the percentage scoring proficient may result in a school being labeled “in need of improvement.” Advocacy groups such as the National Association of Bilingual Educators support the inclusion of LEP students in state testing, arguing that their exclusion from an accountability system is likely to result in decreased educational services. NCLB also requires that LEP students take the state tests in English beginning three years after entering the public school system. The law does not require schools to implement English immersion programs or any other particular approach toward ensuring that students attain proficiency in English.

Despite some positive elements, implementing NCLB is producing profoundly negative effects for LEP students. Over-reliance on standardized tests and punishments combines with lack of funding to frustrate the best efforts of LEP students and the educators working to serve their needs. These problems are becoming somewhat more recognized (Abedi & Dietel, 2004). While some changes have been made in the federal regulations governing NCLB implementation, they have not been remotely adequate to solve the problems.

Problems Defining LEP

Even defining membership in the LEP subgroup is a difficult task because there is no nationally recognized test for English proficiency. States design their own methods for determining exactly who should be labeled LEP. This has long been a challenge for educators of LEP students. NCLB raises the stakes by legislating that schools can be labeled “in need of improvement” based solely on the performance of this subgroup.

Academic English proficiency is defined by experts as “the capacity to use spoken and written English with sufficient complexity that

LEP students have been subject to tests that do not provide valid or reliable results for them.

Over-reliance on standardized tests and punishments combines with lack of funding to frustrate the best efforts of LEP students and the educators working to serve their needs.

one's performance in an academic setting is not impaired" (National Research Council, 2000). Yet, as a student progresses through school, the "academic setting" changes. Therefore proficiency is a moving target governed by the learning context, which is made up of myriad variables, including environment, curricula, teaching styles, and the age of the student.

A single test cannot account for all these factors. This is why many experts urge that proficiency be measured in multiple ways, including oral interviews, teacher checklists, and story retelling (National Research Council, 2000). NCLB requires English proficiency to be measured annually for all LEP students, in addition to math and language arts. Given the relatively high cost of the alternatives, it is likely that standardized tests will be used almost exclusively to determine English proficiency. The results will be highly inaccurate in an area where proficiency is difficult to measure even under ideal circumstances.

The LEP 'Catch-22'

Once a student is classified as LEP and counted toward that group's accountability goal, the issue grows even more complicated. "When [LEP] students improve enough . . . they are taken out of the limited-proficiency category, making it virtually impossible for districts to demonstrate progress" (Dillon, 2003; see also Abedi & Dietel, 2004). This "catch-22" creates a paradox: How is it possible to reach 100 percent proficiency in reading in English by 2014 in a subgroup defined by its limited proficiency in English?

In February 2004, the U.S. Department of Education changed its NCLB regulations to allow LEP students to remain in the subgroup for AYP calculations for up to two years after attaining English language proficiency. This is a step in the right direction, but it does not go far enough. The paradox will remain because the majority of students in the LEP subgroup will continue to be nonproficient *by definition*. Until the unrealistic goal of 100 percent "proficient" on English-language tests is altered for this subgroup, schools and districts with large LEP populations will find themselves at a distinct disadvantage in their pursuit of AYP.

This paradox has forced some states to lower benchmarks for English-language proficiency. Minnesota, for example, only plans to

How is it possible to reach 100 percent proficiency in reading in English by 2014 in a subgroup defined by its limited proficiency in English?

Until the unrealistic goal of 100 percent "proficient" on English-language tests is altered for this subgroup, schools and districts with large LEP populations will find themselves at a distinct disadvantage in their pursuit of AYP.

have 12 percent of LEP students who have been in programs six or more years fully proficient in English by 2013 (Zehr, 2003). California and Illinois have developed plans to give credit to schools in which students progress in learning English. Thus, some state plans simultaneously claim that most of their students with LEP will not be proficient in English, but all of them who have been in the U.S. long enough will reach the “proficient” level on state content exams in English.

Testing Content Areas

The Center for Research on Evaluation, Standards and Student Testing reported that “For the ELL [English Language Learner] student, tests measure both achievement *and* language ability” (Abedi & Dietel, 2004). Testing LEP students in math and English as content areas is just as complicated as assessing English-language proficiency.

As a subgroup, LEP students are counted in AYP calculations and thus are subject to state testing and the accompanying test-induced educational problems. The February 2004 changes loosened NCLB regulations so that states are no longer required to include LEP students’ test scores in AYP calculations during their first year of enrollment in U.S. schools. After the first year, however, LEP students must be counted in AYP calculations for mathematics and language arts even though they will not be English-proficient. Thus, the underlying problem is not solved.

LEP students are permitted certain testing accommodations, including the option to take tests in their native language where available. NCLB allows this for the first three years (and then up to two more years if the local education agency can show special circumstances), but this cut-off point may be too soon. Research shows that academic English proficiency takes, on average, four to seven years to acquire (National Research Council, 2000). Therefore NCLB’s three-year limit on native language testing for LEP students represents one more case of NCLB’s tendency to set students up for failure.

Accommodations are designed to counteract the effects of limited proficiency. Allowing students to take state tests in their native language may enable them to show what they know and are able to do in a context in which their limited grasp of academic English is less of an obstacle. However, if students are taught in English, they may not be

Thus, some state plans simultaneously claim that most of their students with LEP will not be proficient in English, but all of them who have been in the U.S. long enough will reach the “proficient” level on state content exams in English.

Research shows that academic English proficiency takes, on average, four to seven years to acquire.

able to effectively display their knowledge and skills in their native language, while not knowing enough English yet to demonstrate knowledge on English language tests. On a math test, for example, a student may understand the math but be unable to accurately read a word problem in either language.

Since proficiency is so contextual and hard to nail down, it follows that proper accommodations are similarly difficult to standardize or dictate out of context (National Research Council, 2000). Rather than deal with the enormous complexities of designing contextually sensitive accommodations, it becomes easier for schools to simply carry over special education accommodations such as giving students extra time. Research has shown that this is exactly what happens (Peterson, 2002). Furthermore, the U.S. Department of Education's English-language acquisition office is not encouraging states to use alternative tests and is not guaranteeing they will be approved for use (Zehr, 2003b).

Three Cases

Three examples illustrate the negative effect of NCLB on LEP students. Kelly Dawson, a teacher at La Escuela Fratney, a two-way bilingual school in Milwaukee, WI, traces some of her frustrations in a *Rethinking Schools* article (Dawson, 2003). She explains that Wisconsin negotiated a policy with the federal Department of Education that all students with an LEP level of three or above must take the state tests in English. However, the state defines level three and four as students who do not read and write at grade level in English. LEP students were thus being set up for failure.

Dawson goes on to describe the school's preparation for the state testing process. The school went into "panic mode," interrupting their normal curriculum, reassigning special education and mentor teachers from normal duties, and putting in extra hours in a desperate attempt to prepare students to take the state tests in English. Accommodations nearly tripled the total test-taking time, so that LEP students spent 17.5 hours over seven days on the five tests, compared with 6.5 hours for English proficient students. Many of NCLB's flaws – narrowed curriculum, inappropriate tests, undue stress on teachers and students, overtesting – are exacerbated for LEP students.

The second case speaks to the issue of appropriate accommodations. For over a decade, the state of Illinois has been in the process of

On a math test, for example, a student may understand the math but be unable to accurately read a word problem in either language.

The school went into "panic mode," interrupting their normal curriculum, reassigning special education and mentor teachers from normal duties, and putting in extra hours in a desperate attempt to prepare students to take the state tests in English.

developing an English language proficiency test called the Illinois Measure of Annual Growth in English (IMAGE). After several years of extensive field testing and norming, Illinois LEP educators view the test as a reasonably sound instrument for measuring whether students are advancing adequately in English proficiency (Montes, 2003).

Unfortunately, the pressures of NCLB have caused Illinois to use this assessment in a manner that causes it to be harmful to the students and educators it was meant to serve. NCLB requires that states accommodate LEP students “to the extent practicable” (No Child Left Behind Act, 2001). With tight budgets, it is easy for the Illinois State Board of Education to argue that providing state tests in native languages is impracticable. For financial and logistical reasons, the “practicable” course was to use IMAGE as an alternative assessment to the English-language state tests for LEP students. However, IMAGE was never meant to assess learning in relation to the Illinois state standards in language arts or math. The “language arts” portion of IMAGE was created with reference to national English as a Second Language standards, while the math section was added later with minimal field testing. Illinois has fallen into another NCLB trap of using a test for purposes different from those for which it was created.

As Jane Montes, Vice-President of the Illinois Association for Multilingual Education, put it in testimony at public hearings on NCLB: “This entire process has violated basic principles of formal test construction because IMAGE was originally designed as a language test and not as an academic achievement measure . . . In an ideal world, we would recommend that Illinois stop providing meaningless statewide information with regards to the academic achievement of English Language Learners to the U.S. Department of Education until a more valid, reliable, and adequate measure is put into place (Montes, 2003).”

Christie Aird, Director of Elementary Programs for the East Aurora School District in Illinois, testified at the same public hearing as Montes. Aird pointed out that the inappropriate use of IMAGE as a test of academic achievement is contrary to guidelines laid out by the state of Illinois, the American Educational Research Association (AERA), and even NCLB itself. To prove her point, she focused on Rollins Elementary School, which is facing sanctions due to its failure to make AYP for two consecutive years. Although the school posted significant gains in the number of students that met or exceeded state standards in

“This entire process has violated basic principles of formal test construction because IMAGE was originally designed as a language test and not as an academic achievement measure.”

-Jane Montes

The inappropriate use of IMAGE as a test of academic achievement is contrary to guidelines laid out by the state of Illinois, the American Educational Research Association (AERA), and even NCLB itself.

math and English and also scored above the state average in math and writing, the school failed to meet AYP.

This was in large part due to the misuse of IMAGE. The IMAGE proficiency categories of Beginning, Strengthening, Expanding and Transitioning were unscientifically converted to the NCLB categories of Academic Warning, Below Standards, Meeting Standards, and Exceeding Standards for the purposes of the state achievement tests. Based on this conversion, Aird pointed out that one would expect to see IMAGE scores that directly reflected the number of years LEP students had been in the program. At the Rollins School, “That is precisely what happened . . . The students that had 1 to 2 years in the program scored commensurate with the [IMAGE] proficiency categories and received instead [NCLB] ratings that fell into the academic warning and below standards . . . Since scores are weighted equally based on the state formula, the percentage of students taking IMAGE unfortunately caused a decrease in the overall school score” (Aird, 2003).

In other words, the Rollins School may have been doing exactly what is demanded – moving students along the English proficiency continuum at the proper pace – but it was punished because of the state’s misuse of IMAGE. Aird testified that, in effect, a good school was being labeled a failure largely due to NCLB’s mishandling of the LEP subgroup. In response, the Illinois State Board of Education’s liaison for NCLB, Gail Lieberman, was firm: “Every kid has to be tested. And if that’s your population, that’s your population. You need to plan for how you’re going to educate them” (Dell’Angela, 2003). Lieberman simply ignored the possibility that a school may well have been educating its LEP students, but was mislabeled as failing due to the misuse of the IMAGE test to meet AYP requirements.

Conclusion and Recommendations

Overhauling NCLB to avoid harming LEP students and the schools they attend requires the same reforms needed for all students. A high-quality assessment system must replace sole reliance on standardized tests. An irrational, rigid and punitive accountability system must be replaced by an accountability structure that focuses on improvement and provides reasonable flexibility.

The complexities of educating and testing LEP students require that particular attention be paid to ensuring that high-quality assess-

The Rollins School may have been doing exactly what is demanded – moving students along the English proficiency continuum at the proper pace – but it was punished because of the state’s misuse of IMAGE.

An irrational, rigid and punitive accountability system must be replaced by an accountability structure that focuses on improvement and provides reasonable flexibility.

ments are used for the purposes for which they were designed. States should be held accountable for creating strong assessments, not just standardized tests, for English language proficiency and high-quality alternatives to English-language subject tests. States should follow the recommendations of the National Research Council (2000), such as using a variety of appropriate means to assess students' academic and linguistic progress and avoiding inappropriate practices such as arbitrary cut-offs in the length of time English language learners can receive accommodations and other supports. The federal government should provide additional funding for these purposes and establish criteria for success.

The fundamental paradox of LEP testing under NCLB must also be addressed. It is simply unacceptable to continue to tout the goal of 100 percent proficiency for a subgroup defined by its limited proficiency. Rather than focusing merely on the number of students attaining proficiency in academic English, which primarily parallels the amount of time a student has studied the language, schools should be evaluated on the progress that LEP students make on a yearly basis, measured by comprehensive assessments. AYP models for LEP students that acknowledge progress in learning English in addition to the overall percentage achieving English proficiency are being implemented in California and Illinois (Dillon, 2003) and should be expanded and incorporated into NCLB legislation.

In addition to addressing the two main areas of concern to LEP students—inappropriate assessments and AYP irrationality—issues such as adequate funding, teacher quality, and curricula alignment are especially crucial to this vulnerable population. Recommendations in these areas are discussed elsewhere in this report and largely parallel those from expert organizations such as the National Council of La Raza (National Council of La Raza, 2002).

F. Special Education Issues Dog NCLB

As state education officials attempt to comply with NCLB's mandates, the issue of how special needs students fit into AYP and other requirements is coming to the fore.

NCLB includes students with disabilities in the requirement that all students make AYP toward the goal of 100 percent proficiency. Many disability advocates see the inclusion of special needs students

It is simply unacceptable to continue to tout the goal of 100 percent proficiency for a subgroup defined by its limited proficiency.

In addition to addressing the two main areas of concern to LEP students—inappropriate assessments and AYP irrationality—issues such as adequate funding, teacher quality, and curricula alignment are especially crucial to this vulnerable population.

in state tests as the only way to ensure that their educational needs are met and schools held accountable. Others believe the expectation that all students, including those with all but the most severe cognitive disabilities, will score “proficient” by 2014 is an unrealistic hurdle that will doom both students and districts to failure. Still others fear that reliance on standardized tests to determine progress will undermine educators’ ability to respond to students as individuals and may also lead to schools pushing out students with disabilities.

The U.S. Department of Education acknowledged the concern among special educators and in December 2003 released final regulations in an effort to clarify how disabled students should be included in testing (Federal Register, 2003). The regulations permit states to test up to one percent of their total student population, those with the most significant cognitive disabilities, with out-of-grade-level assessments and have their scores counted as “proficient” or “advanced” for meeting AYP. The regulations also allow states to apply to the federal government (and districts to their states) to exceed the one percent cap if they have larger populations of students with the most significant cognitive disabilities. The regulations were greeted with relief in many quarters, at least in part because the law initially did not address the issue of how to handle severely cognitively disabled students. Still, they did not put to rest the controversy over whether NCLB’s approach will be in the best interests of many special needs students’ educations.

In October 2003, the Council for Exceptional Children (CEC) issued a press release on behalf of a group of 150 special educators who expressed grave concerns about NCLB’s potential harm to disabled students (Council for Exceptional Children, 2003). The issues included a concern that state tests may not accurately measure the progress many disabled students are making and that the tests may cause them to feel “humiliated, ashamed, and frustrated.” Contrary to NCLB’s promise to ensure that disabled students achieve at unprecedented levels, the educators fear the law creates pressures that will cause these students to be increasingly excluded from educational opportunities. “The acceptance of students with disabilities is being unraveled,” the press release said. “Students with disabilities are now stigmatized as the ‘group that keeps a school from meeting adequate yearly progress,’ and they are not wanted.”

Others fear that reliance on standardized tests to determine progress will undermine educators’ ability to respond to students as individuals and may also lead to schools pushing out students with disabilities.

The new regulations did not put to rest the controversy over whether NCLB’s approach will be in the best interests of many special needs students’ educations.

Robert Wilson, a 16-year-old from Nashville with attention-deficit disorder, told a *Nashville Tennessean* reporter about his experience with state tests, a description that lends support to the CEC statement. *The Tennessean* reported that Wilson “worries a little about dragging down his school’s overall achievement scores, but mostly he just dreads tests altogether. It’s one of the times he doesn’t feel like every other kid. He feels weird. Out of place. Abnormal. ‘When I am still taking the test and I see people get up, it makes me want to finish up and hurry. It takes me a while to process anything. I hear noises and I look. I feel worried that I’m not going to get all the answers right’” (Riley and Long, 2003).

Wilson said he doesn’t mind spending four or five hours a night on homework and would like to put his artistic talent to work at the Savannah College of Art and Design, but fears he will fail the state tests and be denied a diploma. “I feel really sad, really different and hurt because all the hard work got kicked to the curb,” said Wilson.

Responses

State and local school officials are responding to pressure to improve the scores of disabled students in a variety of ways. Cleveland apparently disregarded the NCLB regulation that a maximum of one percent of all students be assessed based on alternate standards because of severe disabilities. Compared with other districts that used alternate standards for as few as 0.2 percent of disabled students, Cleveland tested 8.8 percent of disabled students using alternate standards and alternate assessments, thereby reaping outstanding results in NCLB report cards released in September (Reed, 2003).

Virginia educational policymakers objected to the same regulation. They argued that this requirement goes against their established policy and violates the federal Individuals with Disabilities Education Act (IDEA). The state’s Standards of Accreditation, adopted in 1997, require that special-education students take Virginia’s regular state tests only if the team developing their individualized education programs (IEP) approves. “We thought that policy was sound when we put it in, and we think it is sound today. We also think it’s required by federal law, the [Individuals with Disabilities Education Act], to let the IEP determine the extent of a [disabled] student’s participation in the SOL [Standards of Learning] program,” said Mark Christie, who was then the president of the Virginia Board of Education (Wermers, 2003).

The Tennessean reported that student Robert Wilson worries a little about dragging down his school’s overall achievement scores, but mostly he just dreads tests altogether.

Wilson fears he will fail the state tests and be denied a diploma. “I feel really sad, really different and hurt because all the hard work got kicked to the curb.”

Virginia is not the only place where conflicts between IDEA and NCLB are being noticed. Knoxville, Tennessee attorney Melinda Baird, a special education lawyer who represents schools, said NCLB and the IDEA “are completely at odds with each other. They’re not reconcilable the way they are written. It sets up an impossible standard for [special needs] kids” (Troutman, 2003).

Baird argued that the two laws are in fundamental conflict because IDEA is individualized and sets up educational programs based on each child’s specific needs, while NCLB “has a one-size-fits-all” standard. She warned that an unintended consequence of NCLB’s AYP mandate could be a reversal of progress toward including more special needs students in regular classrooms.

In an illustration of a potential NCLB/IDEA conflict that is still playing out, Maryland threw out the scores of disabled students who took state tests with accommodations that, according to IDEA, they were legally entitled to because those accommodations are part of students’ IEPs (Mathews, 2003). State education officials believed that the IDEA-mandated accommodations, such as teachers dictating some questions on the third-grade reading section, invalidated the scores for the purposes of determining whether schools and districts had met their AYP targets. As a result, schools had to factor scores of zero for these students into their school averages. Many, therefore, were labeled “in need of improvement.”

Schools have to demonstrate AYP for special education students only when there are enough disabled students in the school to meet or exceed a state-determined “cell size,” which is supposed to be a large enough group to ensure statistically reliable reporting. Minimum cell sizes vary considerably from state to state, ranging from five to 50 students. Reports are circulating that schools are finding ways to reduce the number of special needs students below the threshold in order to avoid falling into the “needs improvement” category. Ironically, rather than holding schools accountable for serving disabled students, this can mean that students who need special education help do not receive it at all. According to University of Florida education researcher Richard Allington, high stakes testing has “put enormous pressure on schools either to better educate the students with disabilities or go figure out how to distribute them in ways such that there are too few to create a ‘testable’ group” (Steiny, 2004).

An unintended consequence of NCLB’s AYP mandate could be a reversal of progress toward including more special needs students in regular classrooms.

Reports are circulating that schools are finding ways to reduce the number of special needs students below the threshold in order to avoid falling into the “needs improvement” category.

Anecdotal reports have surfaced of students getting pushed off the special education rolls for the purpose of improving schools' chances of making AYP targets. One North Carolina mother of a special needs student sits on the board of directors for the state's Exceptional Children's Assistance Center. Sharon Foote says she has heard of "many, many" cases of kids being exited illegally from special ed services to save money and because of the demands placed on schools by NCLB (Foote, 2004). "I have heard it at our meetings. I have heard it from our private advocate who goes to nearly all 120 school buildings in our public school system. I have heard it from the former chair of the school board (whose son is learning disabled). And I have heard it from the special services director at our charter school."

Foote says that schools are taking advantage of the minimum subgroup numbers and trying to shave off enough students so that they fall below the cutoff and therefore do not have to report test results for the students. "Schools are doing all they can to reduce the number of kids receiving [special education] services, so they don't risk being labeled a "failing" school," she said. "And they will even exit kids illegally. They tell the parents the kid made IEP goals and is doing fine now, have the parents sign the paperwork and poof—the kid's no longer LD. And he/she no longer gets the services he/she needs. (But the school looks better.) Isn't that the opposite of leaving no child behind?"

G. More and More Testing

No Child Left Behind increases the stakes attached to tests and mandates a major increase in the amount of state testing. Counting each subject and grade as a separate exam, there were 546 state-mandated tests given to students in the year prior to George W. Bush's signing NCLB in 2001. By the time states have fully implemented current plans for compliance, 1,262 exams will be administered annually for state accountability systems (see Table II – 1 at the end of this chapter), according to research carried out by FairTest in the summer of 2003.

The federal government will require annual testing in reading and math in grades 3-8 and once in high school, plus testing in three grades in science. Two-thirds of the states require or will require testing in subjects and grades beyond the federal mandate. While some of these tests can be attributed to state programs that existed before NCLB,

"They tell the parents the kid made IEP goals and is doing fine now, have the parents sign the paperwork and poof—the kid's no longer LD. And he/she no longer gets the services he/she needs. (But the school looks better.) Isn't that the opposite of leaving no child behind?"

-Sharon Foote

There were 546 state-mandated tests given to students in the year prior to George W. Bush's signing NCLB in 2001. By the time states have fully implemented current plans for compliance, 1,262 exams will be administered annually for state accountability systems.

many more will be added in the overhaul of state testing systems precipitated by the federal mandate. For example, the number of tests in writing and history/social studies, neither of which are required by NCLB, has doubled since states have begun complying with the act. Thus, NCLB seems to be causing a spur to more testing even beyond the federal requirements.

As a result of the expansion of tests in subjects other than reading and math, the distribution of testing amongst grades and subjects has not changed much as a result of NCLB. The focus on areas that were heavily tested prior to the new federal mandates remains essentially the same, with reading and math the most heavily tested subjects, and grades 4 and 8 the years in which students face the most testing. The only notable change in the proportion of testing in each grade has been the growth in grades 6 and 7. These two grades account for twice as much of the testing as they did before NCLB.

In line with NCLB, most states have or are implementing alternatives for students with limited English proficiency. Twenty-two states report plans to provide alternate assessments or translations of state reading and math exams for LEP students. The other approach to assessing English language learners is to use English proficiency exams, some of which are then used for school and district accountability. Only three states — Texas, Michigan, and Pennsylvania — use both alternate assessments or translations and English proficiency exams. This may point towards divergent philosophies on how to assess English language learners within state accountability systems.

Standardized Testing Spreads to the Youngest Students

Among the most troubling aspects of NCLB is its potential to significantly undermine educational quality for young children. Although the testing provisions of the law are not supposed to begin until 3rd grade, in and of itself a young age upon which to impose high-stakes testing consequences, the overemphasis on testing is metastasizing into even earlier grades, with potentially devastating consequences (see Table II – 1 at the end of this chapter).

There is a longstanding consensus among child development experts that younger children be protected from an overemphasis on standardized testing. The Alliance for Childhood advocates for a

NCLB seems to be causing a spur to more testing even beyond the federal requirements.

Among the most troubling aspects of NCLB is its potential to significantly undermine educational quality for young children.

healthy, developmentally appropriate childhood and includes experts in child psychology and early childhood development among its members. The Alliance's statement on high-stakes testing made special reference to the potential harm to young children:

“There is growing evidence that the pressure and anxiety associated with high-stakes testing is unhealthy for children—especially young children—and may undermine the development of positive social relationships and attitudes towards school and learning. A resolution adopted by the National Council of Teachers of English in November 2000 states that ‘high-stakes testing often harms students’ daily experience of learning, displaces more thoughtful and creative curriculum, diminishes the emotional well-being of educators and children, and unfairly damages the life-chances of members of vulnerable groups’” (Alliance for Childhood, 2001).

Reading instruction in the early grades is one area in which NCLB's effect on testing and instruction of young children can be seen. The provisions of NCLB's “Reading First” initiative require assessments to determine whether young students are making satisfactory reading progress. The U.S. Department of Education has in at least some instances mandated the use of standardized tests for this purpose, thereby adding to the total number of tests administered and pushing testing down into lower grades.

For example, North Carolina was recently forced to compromise its own ban against using standardized testing of children in kindergarten through second grade when the federal government said it would not grant the state funds for reading instruction unless it agreed to test those students.

Prior to NCLB, three states gave a total of 12 tests in the K-2 span. More states are responding to NCLB by increasing standardized testing of younger students, with 41 standardized tests in place or proposed in grades K-2 (see Table II – 1 at the end of this chapter). Eight states have now mandated statewide testing in these early grades. Two others, Illinois and Maryland, require testing for only some schools, and one, South Carolina, has optional testing in grades one and two.

Reading instruction in the early grades is one area in which NCLB's effect on testing and instruction of young children can be seen.

For example, North Carolina was recently forced to compromise its own ban against using standardized testing of children in kindergarten through second grade when the federal government said it would not grant the state funds for reading instruction unless it agreed to test those students.

The expansion of testing in the early grades may be attributable in part to NCLB's requirement that each and every school be judged for adequate yearly progress (AYP). This requirement poses a dilemma for schools without grades three and higher, where testing is mandated.

Most states have addressed this dilemma by judging K-2 schools by the grade 3 exam scores of the schools their students go on to attend, but some have addressed the problem by simply adding testing in the earlier grades.

A study of high-stakes accountability systems in Southern states also noted the pressure to prepare for the tests getting pushed down to lower grades. Berry *et al.* (2003) quoted a second grade teacher lamenting the results: "I wonder, do we push more and more down on them? Are we expecting too much? I wonder sometimes, does the school system need to slow down? Let's just sit down and read a book, do some math, instead of trying to cram all this other stuff in. They [the students] get to where they want to give up, 'I can't do any more.' You know, that's not the way we used to teach second grade."

H. Summary

Many of NCLB's assumptions and the model on which it was based have fundamental flaws:

- There is no persuasive evidence that the model of standards, testing, and rewards/punishment for achievement is the cure for what ails public schooling. On the contrary, several studies show a decline in achievement in high-stakes states relative to those with low stakes.
- Surveys of educators and other studies confirm that the model promotes teaching to the test and narrowed curricula, particularly in schools that serve low-income minority students.
- Independent analysts have found that tests often fail to measure the standards and objectives deemed most important by those who have set the standards, meaning that students taught to these tests will not be exposed to high-quality curricula and the public will not be accurately informed about student achievement relative to the standards.
- The quality of instruction tends to suffer under such a model because it is assumed that all students who fail need the same type of instruction. On the contrary, researchers have found that

[The students] get to where they want to give up, 'I can't do any more.' You know, that's not the way we used to teach second grade."

Surveys of educators and other studies confirm that the model promotes teaching to the test and narrowed curricula, particularly in schools that serve low-income minority students.

students fail for different reasons and need different instructional approaches to get on track.

- Research in the U.S. and Britain refutes the assumption that low-achieving students are motivated to work harder and learn more in a high-stakes context. On the contrary, low-achieving students are most likely to become discouraged and give up.
- There is evidence of falling graduation rates in high-stakes states as well as evidence that schools are retaining more students in hopes of reaping higher scores in tested grades. Decades of research support the contention that retained students are more likely to drop out of school permanently.
- The demand that limited English proficient students score proficient on English exams is oxymoronic and puts these students and their teachers in an untenable situation, not conducive to their eventual success in school.
- Special educators and administrators are pleading for more flexibility in determining accountability to recognize the progress and achievement of students who by definition do not learn at the same pace and in the same way as nondisabled students.
- NCLB is fueling a trend toward use of standardized tests in early grades, causing an outcry among early childhood experts who have long cautioned against testing the youngest students this way.

Whether intentional or not, NCLB imposes a separate and unequal education, reduced to test preparation, for the nation's most vulnerable children.

Special educators and administrators are pleading for more flexibility in determining accountability to recognize the progress and achievement of students who by definition do not learn at the same pace and in the same way as nondisabled students.

NCLB imposes a separate and unequal education, reduced to test preparation, for the nation's most vulnerable children.

References

- Abedi, J., and Dietel, R. 2004. "Challenges in the No Child Left Behind Act for English Language Learners," CRESST Policy Brief 7, Winter. Available online at www.cse.ucla.edu/products/newsletters/policybrief7.pdf. See also, Abedi, J. 2004. "The No Child Left Behind Act and English Language Learners: Assessment and Accountability Issues," *Educational Researcher*, V. 33, N. 1, Jan.-Feb. Available online at www.aera.net/pubs/er/pdf/vol33_01/ERv33n1_pp04-14.pdf
- Achieve, Inc. October 2001. "Measuring Up: A Report on Education Standards and Assessments for Massachusetts." Available online at www.achieve.org/achieve.nsf/StandardForm3?openform&parentunid=FCCF5B09747347D285256AE700733993.
- Advocates for Children. 2002. "Pushing Out At-Risk Students: An Analysis of High School Discharge Figures." Available online at <http://www.advocatesforchildren.org/pubs/pushout-11-20-02.doc>.
- Aird, C. September 12, 2003. Testimony for the Education Subcommittee Hearing on the Implementation of the No Child Left Behind Act of 2001. Chicago, Illinois.
- Allensworth, E. April 2004. Ending Social Promotion: Dropout Rates in Chicago after Implementation of the Eighth-Grade Promotion Gate. Consortium on Chicago School Research. Available online at www.consortium-chicago.org/publications/p69.html
- Alliance for Childhood. April 25, 2001. "A Statement of Concern and Call to Action on High-Stakes Testing."
- Amrein, A. and Berliner, D. December 2002. "An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing," Arizona State University. The report is in the peer-reviewed, scholarly journal *Education Policy Analysis Archives* available at epaa.asu.edu/epaa/v10n18/ and the Education Policy Studies Laboratory at ASU.
- AP. October 29, 2003. "Ways to Boost Student Scores Being Considered," *Lexington Herald-Leader*.
- Bernstein, J. August 30, 2002. "Test Case: Hard Lessons from the TAAS," *Texas Observer*.
- Berry, B. *et al.* December 2003. "The Impact of High-Stakes Accountability on Teachers' Professional Development: Evidence from the South," Southeast Center for Teaching Quality Inc., Chapel Hill, NC.

- Bracey, G. 2003. "The Rotten Apples in Education Awards," Education Disinformation Detection and Reporting Agency. Available online at www.america-tomorrow.com/bracey/EDDRA/rottenapples2003.pdf.
- Bryk, T. 2003. "No Child Left Behind, Chicago-Style." In Peterson, P. W., and West, M. *The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press, pp. 242-268.
- Carnoy, M., Loeb, S. Winter 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Education Evaluation and Policy Analysis*, v24(4).
- Center for Inquiry. November 2001. All reports are available online at www.performanceassessment.org/consequences/ccritiques.html.
- Clarke, M. *et al.* November 2002. "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States," National Board on Educational Testing and Public Policy, Boston College. Available online at www.bc.edu/research/nbetpp.
- Coleman, J. S., *et al.* 1966. *Equality of Educational Opportunity*, Washington, DC: United States Department of Health, Education, and Welfare, Office of Education.
- Conley, D. 2003. "Mixed Messages: What State High School Tests Communicate About Student Readiness for College," Center for Educational Policy Research, University of Oregon, Eugene, Oregon.
- Council for Exceptional Children. October 27, 2003. Press Release, "No Child Left Behind Makes 'No Sense' for Students with Disabilities, Say Special Education Teachers." Available online at www.cec.sped.org/PR_NCLB_2003.pdf
- Dawson, K. Summer 2003. "Testing Our Sanity," *Rethinking Schools Online*. Available online at www.rethinkingschools.org/special_reports/bushplan/ESEA174.shtml
- Dell'Angela, T. October 31, 2003. "School Says Exam Leaves Unfair Mark: Controversial Test Brings 'Failing' Tag," *Chicago Tribune*.
- Diamond, J., Spillane, J. November 2002. "The Effects of High Stakes Testing in Low- and High-Performing Schools, Institute for Policy Research, Northwestern University.
- Dillon, S. November 5, 2003. "School Districts Struggle with English Fluency Mandate," *The New York Times*.
- Dobbs, M. December 22, 2003. "School Choice, Limited Options," *The Washington Post*.

- Dobbs, M. April 22, 2004. "'No Child' Law Leaves Schools' Old Ways Behind," *The Washington Post*.
- FairTest *Examiner*. Spring 2002. "Students Criticize TAAS." Available online at www.fairtest.org/examarts/Spring%2002/Students%20Criticize%20TAAS.html.
- FairTest *Examiner*. Fall 2000. "Parents Win Victory in Chicago"; see also, Winter, 1999-2000. "Parents File OCR Complaint." Available online at www.fairtest.org/examarts/Fall%2000/Parents%20win%20chicago.html and www.fairtest.org/examarts/winter00/Parents_File_OCR_Complaint.html.
- Federal Register. December 2003. Title I—Improving the Academic Achievement of the Disadvantaged; Final Rule. Available online at www.ed.gov/legislation/FedRegister/finrule/2003-4/120903a.html.
- Fogg, N. P. *et al.* October 2003. "Youth Labor Market and Education Indicators for the State of Illinois," Center for Labor Market Studies, Northeastern University, Boston, MA.
- Foote, S. February 13, 2004. Personal email correspondence.
- Gayler, K. *et al.* August 2003. "State High School Exit Exams: Put to the Test," Center on Education Policy.
- Haney, W. August 19, 2000. "The Myth of the Texas Miracle in Education," *Education Policy Analysis Archives*. Available online at epaa.asu.edu/epaa/v8n41/.
- Haney, W. 2001. "Revisiting the Myth of the Texas Miracle in Education: Lessons about Dropout Research and Dropout Prevention." Paper prepared for the "Dropout Research: Accurate Counts and Positive Interventions" conference sponsored by Achieve. Available online at www.civilrightsproject.harvard.edu/research/dropouts/haney.pdf.
- Haney, W. September 23, 2003. "Attrition of Students from New York Schools," Invited Testimony at Public Hearing "Regents Learning Standards and High School Graduation Requirements" before the New York Senate Standing Committee on Education. Available online at www.timeoutfromtesting.org/testimonies/923_Testimony_Haney.pdf.
- Haney, W. January 2004. "The Education Pipeline in the United States, 1970-2000," Educational Pipeline Project, National Board on Educational Testing and Public Policy. Available online at www.bc.edu/research/nbetpp/statements/nbr3.pdf.

- Harlen, W., Deakin-Crick, R. June 2002. "A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning," Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre), University of London.
- Hauser, R. and Heubert, J., eds. 1999. "High Stakes: Testing for Tracking, Promotion, and Graduation," Washington, D.C.: National Academy Press.
- Herbert, B. October 20, 2003. "Locked Out at a Young Age," *The New York Times*.
- Hinde, E. 2003. "The Tyranny of the Test," *Current Issues in Education*, cie.asu.edu/volume6/number10/index.html.
- Jacobs, B. Summer 2001. "Getting Tough? The Impact of High School Graduation Exams," *Educational Evaluation and Policy Analysis*, (V.23, N.3).
- Klein, S.P. *et al.* 2000. "What Do Test Scores in Texas Tell Us?" RAND Corp.
- Kober, N. June 2002. "Teaching to the Test: The Good, The Bad, and Who's Responsible," Center on Education Policy.
- Kohn, A. 1999. Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes, *Houghton Mifflin*.
- Koretz, D.M. and Barron, S.I. 1998. "The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)," Santa Monica: RAND.
- Lewin, T. and Medina, J. July 31, 2003. "To Cut Failure Rate, Schools Shed Students," *The New York Times*.
- Linn R. 2000. "Assessments and Accountability," *Educational Researcher* (29,2).
- Mabry, L. *et al.* 2003. "Local Impact of State Testing in Southwest Washington," *Education Policy Analysis Archives*. Available online at epaa.asu.edu/epaa/v11n22
- Mathews, J. October 16, 2003. "Low Test Scores Blamed on Conflicting U.S. Laws," Washington Post. Available online at www.washingtonpost.com/wp-dyn/articles/A30561-2003Oct15.html
- Montes, J. September 12, 2003. Testimony for the Education Subcommittee Hearing on the Implementation of the No Child Left Behind Act of 2001. Chicago, Illinois.

- Moon, T. R., Callahan, C. M., & Tomlinson, C. A. 2003. "Effects of State Testing Programs on Elementary Schools with High Concentrations of Student Poverty - Good News or Bad News," *Current Issues in Education*. Available online at cie.asu.edu/volume6/number8/index.html.
- Moore, D. Winter-Spring 2003. "Research Finds Fault with Chicago's Retention Program," *FairTest Examiner*.
- Nagaoka, J., and Roderick, M. April 2004. "Ending Social Promotion: The effects of retention." Consortium on Chicago School Research. Available online at www.consortium-chicago.org/publications/p70.html.
- National Council of La Raza. 2002. *NCLR Issue Brief*. No. 8.
- National Research Council. 2000. "Testing English Language Learners in U.S. Schools: Report and Workshop Summary," Committee on Educational Excellence and Testing Equity. Kenji Hakuta and Alexandra Beatty, Editors. Board on Testing and Assessment, Center for Education. Washington, D.C.: National Academy Press.
- Peabody, Z. July 24, 2003. "Test Shows HISD 'Achievement Gap' Back," *The Houston Chronicle*.
- Pedulla, J. et al. March 2003. "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers," National Board on Educational Testing and Public Policy, Boston College. Available online at www.bc.edu/research/nbetpp.
- Peterson, B. Fall 2002. "Leaving English Learners Behind," *Rethinking Schools*. Available online at www.rethinkingschools.org/special_reports/bilingual/Leav171.shtml.
- Reed, E. September 16, 2003. "Cleveland schools take advantage of test loophole for disabled," *The Plain Dealer*.
- Riddle Buly, Marsha and Valencia, Sheila W. Fall 2002. "Below the Bar: Profiles of Students who Fail State Reading Assessments," *Educational Evaluation and Policy Analysis*, depts.washington.edu/ctpmail/PDFs/Reading-MRBSV-04-2003.pdf.
- Riley, C., Long, D. October 29, 2003. "Special ed dilemma," *The Tennessean*. Available online at www.tennessean.com/education/special/nclb/archives/03/10/41541677.shtml
- Rose, L. C. and Gallup, A. M. September 2003. "The 35th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward The Public Schools," *Phi Delta Kappan*.

- Rosenfeld, E. February 22, 2004. "The Best Answer," *Washington Post Sunday Magazine*.
- Rothman, R. *et al.* May 2002. "Benchmarking and Alignment of Standards and Testing," CRESST.
- Ryan, R. M., & La Guardia, J. G. 1999. "Achievement Motivation within a Pressured Society: Intrinsic and Extrinsic Motivations to Learn and the Politics of School Reform." In T. Urdan (Ed.) *Advances in motivation and achievement: Vol 11*. Greenwich, CT: JAI Press.
- Schemo, D. August 28, 2003. "For Houston Schools, College Claims Exceed Reality," *The New York Times*.
- Shanklin, M. March 5, 2004. "School cuts students; grade goes up," *Orlando Sentinel*.
- Shepard, L. A. and Smith, M. L. 1989. "Flunking Grades: Research and Policies on Retention," Philadelphia: Falmer.
- Stecher, B. *et al.* November 2003. "*Working Smarter to Leave No Child Behind: Practical Insights for School Leaders*," RAND Corp.
- Steiny, J. January 25, 2004. "Gaming school accountability," *Providence Journal*.
- Swanson, C. B. May 2003. "Caps, Gowns, and Games: High School Graduates and NCLB," Urban Institute.
- Troutman, N., May 2, 2003, "Problems with IDEA," *Nashville City Paper*.
- WCER. Fall 1999. "Are State-Level Standards and Assessments Aligned?" WCER Highlights, 1-3. Madison: Wisconsin Center for Educational Research.
- Wermers, J. April 30, 2003. "Va., U.S. Differ on Tests for Special-Ed Students," *Richmond Times-Dispatch*.
- Werner, A. November 20, 2003. "HISD: Cheating kids out of an education," KHOU.
- von Zastrow, C. March 2004. "Academic Atrophy: The Condition of the Liberal Arts in America's Public Schools," Council for Basic Education.
- Zehr, M. November 19, 2003. "English Proficiency Can Take a While in State ESEA Plans," *Education Week*.
- Zehr, M. November 26, 2003. "States Developing Tests for English-Learners," *Education Week*.

Table II – 1: Total number of state exams projected for 2014

	ELA	Reading	Writing	Math	Science	Soc. St.	Other	Total by Grade
K	0	3	2	3	0	0	0	8
1	2	4	2	4	0	0	0	12
2	5	7	2	7	0	0	0	21
3	18	39	6	52	14	9	0	138
4	20	40	14	52	38	12	2	178
5	20	38	8	52	23	12	4	157
6	19	39	6	52	12	9	0	137
7	20	38	11	52	16	9	0	146
8	19	40	10	52	43	16	4	184
9	7	7	2	11	4	1	1	33
10	9	15	9	21	13	7	1	75
11	8	10	5	16	13	7	1	60
12	2	1	0	2	3	2	0	10
HS	11	12	1	26	30	19	4	103
Total by Subject	160	293	78	402	209	103	17	<u>1262</u>

Table II-2: Tests Added Summary

	ELA	Reading	Writing	Math	Science	Soc St.	
K	0	3	2	3	0	0	8
1	1	3	2	3	0	0	9
2	2	4	2	4	0	0	12
3	14	21	4	32	8	6	85
4	11	21	3	26	30	7	98
5	18	20	5	32	15	6	96
6	16	27	3	39	8	6	99
7	19	30	5	44	10	7	115
8	6	13	4	13	33	7	76
9	6	4	1	8	2	1	22
10	3	4	1	7	7	4	26
11	5	3	2	7	5	3	25
12	2	0	0	1	3	1	7
HS	3	3	0	9	24	5	44
	106	156	34	228	145	53	<u>722</u>