
Failing Our Children

• How “No Child Left Behind”
Undermines Quality and Equity in Education

• An Accountability Model That
Supports School Improvement



by Monty Neill, Lisa Guisbond and Bob Schaeffer
with James Madden and Life Legeros
May 2004

FairTest

The National Center for Fair and Open Testing

Failing Our Children

• How “No Child Left Behind”
Undermines Quality and Equity in Education

• An Accountability Model That
Supports School Improvement

by Monty Neill, Ed.D., Lisa Guisbond and Bob Schaeffer
with James Madden and Life Legeros
May 2004
\$30

FairTest

The National Center for Fair and Open Testing

FairTest: The National Center for Fair & Open Testing
342 Broadway
Cambridge, MA 02139
(617) 864 - 4810; fax (617) 497-2224
www.fairtest.org
email: info@fairtest.org

Thank you to the Joyce Foundation for grant support in the preparation of this report, and to FairTest's many friends and donors who make our work possible. None are responsible for the contents of this report.

Table of Contents

Executive Summary	ES-1
Introduction. No Child Left Behind Testing and Sanctions Provisions	i
Chapter I. Set Up to Fail	1
Appendix 1. Jeb’s A+ Schools Dubbed Failing by George	23
Appendix 2. When AYP Means Good Schools Are Called Bad	24
Chapter II. Testing: What’s the Problem with ‘No Child Left Untested’?	35
Chapter III. False Promises & Flawed Policies: Navigating the NCLB Maze	83
Chapter IV. Money, Education and Accountability	120
Chapter V. Organizing for Change Amid Growing Opposition to NCLB	131
Appendix 1. What’s in a Name? “Renaming” NCLB	143
Chapter VI. Authentic Accountability	146

Tables and Charts

Table I - 1. Data on Schools Not Making Adequate Yearly Progress (AYP)	30
Table I - 2. State AYP Starting Points	32
Table I - 3. State Cell Sizes: Minimum Number of Students to Be in a Cell	33
Chart I - 1. Cell Sizes: Number of States Selecting Given Cell Size	34
Table II - 1. State Tests Projected for 2014.	82
Table II - 2. State Tests to Be Added by 2014	82

© 2004 FairTest.

Permission is given to reproduce this document, or portions of it, except that permission must be sought for commercial use of content from this publication, when the material quoted is used in advertising, or when charges for copies beyond the cost of reproduction are made.

Failing Our Children

• How “No Child Left Behind”
Undermines Quality and Equity in Education

• An Accountability Model That
Supports School Improvement

by Monty Neill, Ed.D., Lisa Guisbond and Bob Schaeffer
with James Madden and Life Legeros
May 2004
\$10

FairTest

The National Center for Fair and Open Testing

FairTest: The National Center for Fair & Open Testing
342 Broadway
Cambridge, MA 02139
(617) 864-4810; fax (617) 497-2224
www.fairtest.org
email: info@fairtest.org

Thank you to the Joyce Foundation for grant support in the preparation of this report, and to FairTest's many friends and donors who make our work possible. None are responsible for the contents of this report.

© 2004 FairTest.

Permission is given to reproduce this document, or portions of it, except that permission must be sought for commercial use of content from this publication, when the material quoted is used in advertising, or when charges for copies beyond the cost of reproduction are made.

Failing Our Children:

- How “No Child Left Behind” Undermines Quality and Equity in Education
- An Accountability Model That Supports School Improvement

Executive Summary

“No Child Left Behind” (NCLB), the title of the federal Elementary and Secondary Education Act, describes a worthy goal for our nation. Tragically, NCLB is aggravating, not solving, the real problems that cause many children to be left behind. NCLB must be overhauled if the federal government is to make a useful contribution to enhancing the quality of education received by low-income and minority group students.

Prior reports and articles on NCLB have exposed a host of problems. Most of these studies have focused on the effects of NCLB. A few have sought to expose the law’s conceptual and structural flaws. Some have proposed modest changes, but only minor regulatory revisions have been adopted. None of these studies have proposed a comprehensive, alternative approach to the federal role in improving public schools for students in poverty.

This report details the fundamental errors in the conception and execution of NCLB. Beyond that, it proposes a better system of accountability to serve as the basis for a comprehensive overhaul of NCLB.

Two false assumptions undergird NCLB:

1) *Boosting standardized test scores should be the primary goal of schools.* This assumption leads to one-size-fits-all teaching aimed primarily at test preparation, and it works against efforts to give all children a high-quality education.

2) *Schools can best be improved by threatening educators with harsh*

This report details the fundamental errors in the conception and execution of NCLB. Beyond that, it proposes a better system of accountability to serve as the basis for a comprehensive overhaul of NCLB.

sanctions, since poor teaching is the primary cause of unsatisfactory student performance. Threats may get teachers to focus narrowly on boosting test scores. They fail, however, to address the underlying problems of family poverty and inadequate school funding that are major reasons why many students start off far behind and never catch up.

A new accountability system must start from accurate assumptions, which will lead to a fundamentally different approach than NCLB's test-and-punish methodology. This alternative approach assumes educators want to do their jobs but often need assistance to do better. Rather than threaten educators with sanctions based on limited information provided by test scores, this alternative approach focuses on gathering multiple forms of evidence about schools and then using that evidence to support comprehensive improvements in teaching and learning. All levels of government must help schools build the capacity to ensure all children receive a high-quality education that meets their individual needs. Governments, therefore, must fulfill their responsibilities to provide adequate and equitable resources. This alternative approach also encourages parents and the community to be core participants in keeping local schools accountable, rather than excluding them through incomprehensible statistical procedures dictated by remote bureaucrats.

What makes NCLB so dangerous is the way it links standardized testing with heavy sanctions through the rigid "adequate yearly progress" (AYP) formula. Thus, the weaknesses of standardized exams – their cultural biases and their failure to measure higher order thinking – are reinforced by strict penalties. The consequence of narrow exams and strong sanctions is intensive teaching to the test. This response undermines decent education as well as efforts to ensure genuine improvements in educational quality.

The lack of adequate funding for schools and for the well-being of children intensifies these problems for low-income and minority-group students. Overcrowded classrooms make it more likely that teachers focus on little more than the content of mandatory tests. The convergence of testing, sanctions and inadequate funding means too many children will continue to get a second-class education. A false accountability system based on testing and punishing will never bring about success for all children. This goal will be out of reach as long as there is

A new accountability system must start from accurate assumptions, which will lead to a fundamentally different approach than NCLB's test-and-punish methodology.

What makes NCLB so dangerous is the way it links standardized testing with heavy sanctions through the rigid "adequate yearly progress" (AYP) formula.

worsening poverty and inadequate funding that denies too many students access to a rich and comprehensive curriculum to prepare them to be lifelong learners, active participants in our democracy and successful in further education and employment.

Despite some potentially helpful provisions, the deep flaws in NCLB end up undermining both educational quality and equity. For example:

- NCLB calls for multiple measures that assess higher order thinking and are diagnostically useful. However, these provisions are not enforced by the U.S. Department of Education and are not embedded in most state practices. The push for standardization and the requirements for quickly imposing “in need of improvement” judgments and sanctions make it nearly impossible for states to implement an assessment system that fosters high-quality learning.
- The law mandates school (or district) improvement plans. In practical terms, however, “improvement” means no more than boosting test scores. Quick sanctions based on unrealistic rates of “adequate yearly progress” deny schools the opportunity to see if their improvement efforts will work.
- The law also requires schools “in need of improvement” to use some of their federal money for professional development. This is reasonable, and the law allows many varieties of professional development to be pursued. But funding is inadequate, and the power of testing will tend to reduce professional development to training teachers to be better standardized exam coaches.

This report details other failures of NCLB:

- The gauge of student progress in most states is being reduced to reading and math scores. Many schools already are narrowing instruction to what is tested.
- Most schools will fail to meet the unrealistic demands imposed by the law’s “adequate yearly progress” provision. Virtually no schools serving low-income children will clear the arbitrary hurdles. Many successful schools are being set up to fail and will be forced to drop policies and programs that work.

The gauge of student progress in most states is being reduced to reading and math scores.

Most schools will fail to meet the unrealistic demands imposed by the law’s “adequate yearly progress” provision.

- Sanctions intended to force school improvement will do the opposite. Because the sanctions merely shuffle inadequate resources, they will pit parents against teachers, parents against parents from other schools, and schools against schools. They divert funding from helping all children succeed to helping the relatively few whose parents want and can obtain transfers and tutoring, and to the businesses that provide these services. The law's ultimate sanctions – privatizing school management, firing staff, state takeovers, and similar measures – have no proven record of success.
- The federal government has failed to adequately fund the law. Education resources are dwindling in most states just as they are hit with the demands of the current law. Worse, neither federal nor state governments are addressing either the resources required to bring all children to educational proficiency or the deepening poverty that makes it difficult for so many children to learn.

These problems have catalyzed a growing movement seeking to overhaul NCLB. From state officials to local parents and teachers, people across the nation are mobilizing against the law. The sources of the anger vary, as do the proposed solutions, but a critical mass for change is building. Unfortunately, some efforts to minimize the damage caused by NCLB would perpetuate educational inequality. Others address only peripheral issues, not the law's faulty assumptions.

Effective opposition to NCLB must embrace genuine accountability, stronger equity, and steps to really improve schools. What is needed is a law founded on a more realistic understanding of the problems facing schools and the processes of school change. A new NCLB must propose solutions that improve, not undermine, the quality of education.

Core elements of a better accountability system include:

- 1) Federal, state and local governments must work together to provide a fair opportunity for all children to learn a rich curriculum in a supportive yet challenging environment. Governments have generally failed to meet this fundamental accountability requirement because they have not ensured

Sanctions intended to force school improvement will do the opposite.

The federal government has failed to adequately fund the law.

These problems have catalyzed a growing movement seeking to overhaul NCLB.

Effective opposition to NCLB must embrace genuine accountability, stronger equity, and steps to really improve schools.

adequate, equitable funding and because they have primarily emphasized test scores.

- 2) Accountability systems must use multiple forms of evidence of student learning. If we want to know how well students are doing, we need to look at a range of real student work. If we want students to learn more or better, we have to provide teachers and students with useful feedback based on high-quality classroom assessments that encompass a variety of ways to demonstrate knowledge and that fit with how children really learn.
- 3) Accountability systems must focus on helping teachers and schools ensure educational success for all students. They must also ensure that schools are safe, healthy, supportive and challenging environments. This means providing data that can be used for improvement and making certain that schools are well equipped to use it. It also requires ample time and resources to enable teachers to learn more, share knowledge and get better at what they do.
- 4) Accountability systems must involve those most directly affected and closest to the classroom. Therefore, the primary accountability mechanisms must be local. They must involve educators, parents, students and the local community; and they must use participatory processes such as local school councils, annual reports and meetings to review school progress.
- 5) The primary responsibility of state governments is to provide tools and support for schools and teachers to improve while ensuring that equity and civil rights are maintained. Intervention should take place only when localities have been given resources and support and still fail to improve, or when there are uncorrected civil rights violations.

In the short term, NCLB should be amended to stop the destructive inflexibility of the “adequate yearly progress” provisions and eliminate the draconian penalties. The requirement for states to annually test all students in grades 3 to 8 in reading and math should be removed and the amount of required testing reduced. Additional

Accountability systems must use multiple forms of evidence of student learning.

Intervention should take place only when localities have been given resources and support and still fail to improve, or when there are uncorrected civil rights violations.

measures of school and student learning should be included in progress evaluations. Congress also must appropriate the full amount authorized under NCLB.

More fundamentally, policy-makers must consider the damage NCLB has wrought and take seriously the problem of inadequate educational funding around the nation. They should begin by listening to the rising voices of educators, parents and community people who know that their children cannot be reduced to test scores. Children deserve a high-quality education, not test preparation.

In its current form, NCLB is a punitive law that uses flawed standardized tests to falsely label many schools as failures, then punish them with harmful sanctions. NCLB must be transformed into a supportive law that promotes lasting educational improvement and makes good on the promise, in the words of the Children's Defense Fund, to "leave no child behind."

NCLB must be transformed into a supportive law that promotes lasting educational improvement and makes good on the promise, in the words of the Children's Defense Fund, to "leave no child behind."

Introduction: No Child Left Behind Testing and Sanctions Provisions

No Child Left Behind (NCLB), the 2001 reauthorization of the federal Elementary and Secondary Education Act (ESEA), includes requirements for state-level testing, mandates for schools and districts to demonstrate “adequate yearly progress” (AYP), and sanctions to be imposed for failure to make AYP. The provisions are in Title I, the section of the law intended to improve education for “disadvantaged” children.

Testing

Under NCLB, each state is required to immediately adopt content and performance standards in both reading and math, with assessments based on these standards. Science requirements will be added later this decade. Standards and tests must establish three levels of academic performance — advanced, proficient, and basic.

Currently, states must assess every student in reading and math at least once in grades 3-5, 6-9 and 10-12. By the 2005-06 academic year, states must assess each child every year in grades 3-8 and once in grades 10-12 in math and reading. By 2007-08, states must add an annual science assessment in the three grade spans.

NCLB permits state assessment systems that are a mix of state-mandated exams and local assessments, but almost every state is relying solely on state tests. While tests are supposed to be based on state-approved standards, U.S. Department of Education regulations allow the use of commercial norm-referenced tests, which may not be standards-based.

The law requires all assessments to be valid and reliable for the purposes for which they are used and “involve multiple, up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills.” The assessments also must “produce individual student interpretive, descriptive, and diagnostic reports...that allow parents, teachers and principals to understand and address the specific needs of students.”

By the 2005-06 academic year, states must assess each child every year in grades 3-8 and once in grades 10-12 in math and reading.

While tests are supposed to be based on state-approved standards, U.S. Department of Education regulations allow the use of commercial norm-referenced tests, which may not be standards-based.

Testing accommodations and alternative assessments must be available for students with disabilities. Limited English proficient students are to be assessed “where possible” in “language and form most likely to yield accurate data” on student achievement, until they reach proficiency in English; and to be assessed annually for English proficiency if not yet proficient.

Accountability and “Adequate Yearly Progress” (AYP)

States must have a “single statewide accountability system” to ensure that districts and schools make “adequate yearly progress” (AYP). The accountability process began with data from the 2001-02 school year, except for schools that were already under improvement or correction status. Within 12 years (by 2014), all students are supposed to attain the proficient level on the state assessments in reading and math. “All students” is defined to exclude those with severe special needs and those who have not been in U.S. schools for three years and who have limited English proficient status.

A school or district can fail to make AYP if a sufficient percentage of all its students or its students in a defined group – low-income, limited English proficient, racial/ethnic minority (African American, Hispanic, Asian, Native American Indian), or students with disabilities – do not reach the proficient level in either math or reading, or if it does not test 95 percent of its students in any category. To be separately counted as a group, there must be enough students to measure reliably. States have set those minimum “cell” levels as low as five students, but most are around 30 (see Chart I-1 at the end of Chapter I).

States determine their starting points based on formulas in the law. The gap between the percentage of students who are at the starting point and 100 percent proficient must be closed by an average of 1/12 per year. For example, if the state’s starting point is 40 percent, then the number of students at the proficient level must increase by an average of 5 percentage points per year.

A state can use multiyear averages to set AYP goals. If it does, the first proficiency target must be in two years, and subsequent targets must be set at intervals of no more than three years. States can start with a slow rate of increase and speed up toward the end; some states

Within 12 years (by 2014), all students are supposed to attain the proficient level on the state assessments in reading and math.

Schools that don’t make AYP but reduce the percentage of students not reaching proficient by at least 10 percent can avoid sanctions under this provision.

expect to make about half their required progress in the last few years leading up to 2014. States also can apply more complex formulas to define rates of progress.

Some schools will start out with a smaller percentage of proficient students than the overall state starting point. Those schools must catch up and make AYP. However, there is a “safe harbor” provision in the law. Schools that don’t make AYP but reduce the percentage of students not reaching proficient by at least 10 percent can avoid sanctions under this provision.

Sanctions

While all schools in a state must report test results, only schools receiving NCLB Title I funds are subject to sanctions. (Nationally, 95 percent of school districts and 50 percent of schools get Title I funds.) Districts will “identify for school improvement” any school receiving federal Title I funds that fails for two consecutive years to make AYP, and states will similarly identify districts. An identified school (or district) must craft an improvement plan listing “specific, measurable objectives” for ensuring all students in each group make AYP.

Students at schools so identified will be eligible to transfer to another public school in the district that is not “in need of improvement” (INOI). This can be another school with similar scores that does not receive Title I money.

If a school does not make AYP after one year under its improvement plan, the district must make tutoring available. This can include parent choice among district-approved private companies or public schools that are not INOI. Districts are required to spend up to 20 percent of their Title I funds for transportation or tutoring programs.

If after the second year of the improvement plan the school or district is not making AYP, it will be identified for “corrective action.” The identified school or district must implement options from a list specified in the law. If after one full year of corrective action a school or district still does not make AYP, it must undergo “restructuring.”

If a school does not make AYP after one year under its improvement plan, the district must make tutoring available.

At the school level, “restructuring” is to include at least one of the following changes in governance:

- Reopen the school as a “public charter school”
- Replace “all or most of the school staff”
- Enter into a contract “with an entity, such as a private management company...to operate the school”
- Turn “the operation of the school over to the state” or
- “Any other major restructuring of the school’s governance arrangement that makes fundamental reform.”

States are to apply a similar list of sanctions to failing districts.

If a school (or district) identified for improvement, corrective action, or restructuring makes AYP for two consecutive school years, requirements for improvement are suspended.

References

NCLB law, regulations, guidance and promotional materials are online at <http://www.ed.gov/policy/landing.jhtml?src=ln>.

FairTest fact sheets, articles and links to reports on NCLB are at <http://www.fairtest.org/nattest/bushtest.html>.

“It’s amazing to me how ridiculous this is. It’s almost as if everybody has been set up to fail.”

– Brenda Montoya, Las Vegas parent

“The ESEA [No Child Left Behind Act] is like a Russian novel. That’s because it’s long, it’s complicated, and in the end, everybody gets killed.”

— Scott Howard, former superintendent, Perry, Ohio, public schools

I. Set up to Fail

As states and districts tally test results and compile long lists of schools that have failed to make “adequate yearly progress” (AYP), the prospect of most U.S. public schools facing sanctions under the federal No Child Left Behind (NCLB) law seems increasingly inevitable. In just the first two years of this marathon race toward 100 percent proficiency, one quarter of all U.S. schools have already failed to make AYP (Center on Education Policy, 2004). Florida had the most schools on the failing list, with 88 percent (Miller, 2003). But Alaska, Delaware, Missouri, North Carolina and Oregon all have had 50 percent or more of their schools labeled as failing to meet targets for improvement (eSchool News, 2003).

Whether or not NCLB’s authors intended to set up the vast majority of public schools for failure, state takeover and possible private management, most observers agree that, barring substantive changes to the current law, this will be the likely result of the requirement that all students score “proficient” on state tests by 2014.

Not surprisingly, a large and growing number of those whose schools have landed on these AYP warning lists are criticizing NCLB for a range of defects. Families whose children attend public schools are receiving confusing and contradictory messages about their schools, rather than the clear and useful information promised by President Bush. In many cases, Florida being a prime example, schools showed consistent and marked improvement in state rankings, yet were judged to have failed when subjected to the NCLB formula.

The public at large (those with no direct involvement in public schools) is also seeing a somewhat confusing but predominantly negative portrait of public education. Daily newspapers report long lists of

In just the first two years of this marathon race toward 100 percent proficiency, one quarter of all U.S. schools have already failed to make AYP.

Families whose children attend public schools are receiving confusing and contradictory messages about their schools, rather than the clear and useful information promised by President Bush.

“failing” public schools. They also feature school officials disputing the results, expressing exasperation, or pleading for a more balanced and nuanced view of how the schools are doing—not to mention more money to run their schools and administer the tests in a time of fiscal austerity.

Harvard Graduate School of Education Professor Richard Elmore is among those who say that NCLB’s AYP provision is ungrounded in any proven theory of how schools actually improve. “The process of genuine improvement does not occur in equal annual increments. The AYP requirement, a completely arbitrary mathematical function grounded in no defensible knowledge or theory of school improvement, could, and probably will, result in penalizing and closing schools that are actually experts in school improvement” (Elmore, 2003).

Rather than provide substantive answers to the many questions raised about NCLB, the U.S. Department of Education (DOE) launched a major public relations campaign to counter NCLB “negativism.” The DOE allocated \$500,000 to assemble a team to run the No Child Left Behind “Communications and Outreach” operation. The group, headed by a former Bush campaign operative, was tasked with supplementing existing DOE communications work. As criticism of the law reached a crescendo, Education Secretary Rod Paige and President Bush himself went on a public relations blitz to repair NCLB’s flagging reputation. But no PR campaign can compensate for the deep flaws in the law, nor for the lack of adequate resources for struggling schools.

In a democratic system that depends on the contributions of all its citizens to the funding of public schools, NCLB’s inaccurate picture of widespread, consistent failure is itself a major threat to the future of public education. The underlying structural dynamic of NCLB produces a vastly distorted picture, tarring both successful and underperforming schools as failures. Because of this distortion, NCLB risks not only undermining crucial support for public schools, but making it impossible to determine which schools and individual students really need substantial support and/or guidance on how to foster academic success.

This chapter will look at the predictions of massive failure to meet NCLB’s targets for school improvement, show how those predictions

“The AYP requirement, a completely arbitrary mathematical function grounded in no defensible knowledge or theory of school improvement, could, and probably will, result in penalizing and closing schools that are actually experts in school improvement”

-Richard Elmore

NCLB risks not only undermining crucial support for public schools, but making it impossible to determine which schools and individual students really need substantial support and/or guidance on how to foster academic success.

have been borne out, explain how such failure was built into the design of NCLB, and then illustrate what's wrong with relying on test scores alone to assess and improve public education.

A. The AYP Mess

The Adequate Yearly Progress (AYP) provision has many flaws:

- Widespread failure of schools to meet AYP targets was predictable and was indeed predicted by many experts who analyzed the law's provisions.
- After just two years, predictions of failure have been borne out in long lists of schools and districts landing on watch lists and lists of schools "in need of improvement."
- There are so many ways to fail under NCLB that it is difficult to draw comparisons between one failing school and another.
- High-poverty schools and districts are overwhelmingly the first to be identified as failing to make AYP.
- Diversity itself is penalized by the AYP formula. The more subgroups a school has, the less likely it will be to make AYP.
- Even well-off suburbs are not immune from failure if their schools include groups of students that struggle to perform well on state tests.

Widespread Failure: Predicted and Predictable

Once analysts began to digest NCLB's intricate provisions, it became clear even before the bill passed that high rates of school failure were the logical outcome of the law's approach to assessment and accountability. David Shreve, of the National Conference of State Legislatures, reflected the consensus of researchers when he estimated that 70 percent of all schools would be labeled "in need of improvement" in the coming years (Prah, 2002). State projections varied based on the difficulty of state tests, the rate of improvement expected, and the size of the subgroup chosen by the state as "statistically significant," but many projected massive levels of failure.

California's prospects for failure were perhaps most extreme. Under its performance standards, 98 percent of all schools in the state and 100 percent of schools serving mostly low-income students were expected to fail to meet the AYP goal. State Education Secretary Kerry Mazzone explained, "We would rather set the bar high and not have everyone reach it than set it low and have everyone

David Shreve, of the National Conference of State Legislatures, reflected the consensus of researchers when he estimated that 70 percent of all schools would be labeled "in need of improvement" in the coming years.

State Education Secretary Kerry Mazzone explained, "We would rather set the bar high and not have everyone reach it than set it low and have everyone reach it"

reach it” (Helfand, 2003). California’s AYP plan required 7 percent per year gains, but in 2002 the state’s actual test score gain was only about 1.5 percent. In a July 2003 report, the state Department of Education said just 32 percent of California schools achieved adequate progress for the year, lending support to dire predictions. Other states expecting over 90 percent of their schools to “fail” include Maine and Massachusetts, with Louisiana projecting 85 percent (Maine Education Association, 2003).

In 2002, a group of researchers met to discuss the AYP formula and to predict outcomes. Edward Haertel, of Stanford University, noted that if progress were based on the experience of the National Assessment of Educational Progress (NAEP) test results over the years, it would take 110 years to reach 100 percent proficiency across the country (Linn *et al.*, 2002c).

Some proponents of NCLB argue that the law’s “safe harbor” provisions will give schools breathing room. “Safe harbor” applies to subgroups that do not make AYP, if the percentage of students in that group decreases by ten percent from the previous year and that group made progress on another academic indicator. Others have pointed to the use of “rolling averages” (i.e., averaging scores over several years) as a means of reducing the impact of not doing well in one year. However, an analysis of state scores in Maine and Kentucky from the 1990s found that rolling averages will have very little impact, and safe harbor only a modest impact. Researcher Jaekyung Lee (2004) concluded, “Contrary to some expectations, the applications of both options would do little to reduce the risk of massive school failure due to unreasonably high AYP targets for all student groups.”

INOI Lists Confirm Dire Predictions

Any hopes that predictions of large-scale failures were inaccurate or exaggerated were dashed by state-by-state lists of schools dubbed “in need of improvement” (INOI) based on state test results from the 2002-2003 school year (see Table I-1 at end of chapter). If anything, predictions underestimated the extent of the failures and the confusion caused by constantly changing lists of “failing” schools. A national teachers union estimated that 26,000 of the nation’s 93,000 public schools failed to make adequate yearly progress in 2004.

- In New York City, 40 percent of the schools were labeled failing (Gootman, 2003).

Edward Haertel, of Stanford University, noted that if progress were based on the experience of the National Assessment of Educational Progress (NAEP) test results over the years, it would take 110 years to reach 100 percent proficiency

If anything, predictions underestimated the extent of the failures and the confusion caused by constantly changing lists of “failing” schools.

- In New Mexico, more than 70 percent of the schools statewide would have failed if the new standards had been applied, so it got an extension of its compliance deadline from the U.S. DOE (Hutton, 2003).
- Fifty-seven percent of Delaware's public schools failed to make adequate yearly progress in math and reading, with 25 of the state's 28 high schools rated as under academic review, including one school that *Newsweek* magazine recently touted as among the best in the nation. At the middle school level, only three of the state's 33 middle schools made adequate yearly progress (Fuetsch, August 12).
- Just 32 percent of California schools achieved adequate progress. Elementary schools fared the best, with 37.2 percent meeting the benchmarks, but the achievement rate dropped to 19.7 percent in middle and high schools.
- In Mississippi, Pascagoula Schools Superintendent Hank Bounds said the new AYP guidelines may mean every school in the district might be classified as failing by the federal government — even the highest performing ones.
- In Michigan, 896 of the state's 3,472 public schools did not make AYP (Jacques, 2004).
- Lists of failing schools turned out to be fluid in Illinois, Texas and Minnesota. After an appeals process, Minnesota moved nearly half the schools, 93, off the failing list (School Funding Services, 2004).

“Failure lists” are in flux for a number of reasons. The unprecedented flood of data is bound to include human and other errors, so that some schools are erroneously included or excluded. There is also the temptation on the part of schools, districts, and even state education agencies to underreport numbers of failing schools, or at least spin the numbers in the most positive way.

A Massachusetts Department of Education press release, for example, emphasized statewide results showing that 94 percent of students made AYP (Massachusetts DOE, 2003). This looked much more positive than the fact that 67 percent of districts in the state were failing to make AYP because of the performance of one or more subgroups and were therefore on the path to possible NCLB sanctions (for a list of sanctions, see “Introduction: No Child Left Behind Testing and Sanctions Provisions”). The DOE press release did include the 67 percent statistic, but buried it near the bottom of the page.

In Mississippi, Pascagoula Schools Superintendent Hank Bounds said the new AYP guidelines may mean every school in the district might be classified as failing by the federal government — even the highest performing ones.

“Failure lists” are in flux for a number of reasons. The unprecedented flood of data is bound to include human and other errors, so that some schools are erroneously included or excluded.

Failure Comes in All Shapes and Sizes

Beyond sharing the stigma of failure and the risk of sanctions, many schools failing to meet AYP targets have very little in common with one another. Some schools clearly are failing to provide what their students need to be successful in higher education, life and work, although for a range of different reasons. Many are making improvements and progress, but not at a rate considered fast enough. Some have limited resources but nevertheless offer good educations to students who come to schools with enormous and growing needs. The vast majority of the schools with the farthest to go are in high-poverty urban communities. But NCLB also fails suburban schools rich in resources that comply with all but one or two of the law's many mandates.

High-Poverty Schools Fail First

To virtually no one's surprise, high-poverty urban school districts are characterized by high, in some cases nearly universal, failure to make AYP. A report released in July 2003 by Michigan State University's Education Policy Center, for example, found, "Nearly all of Michigan's most troubled schools are in high-poverty urban areas and serve low-income, minority children." The report found that only seven out of the 216 troubled schools were in suburban and rural areas. David Plank, director of the center, said it is no secret that the poorest city schools perform most poorly, but the proportion—97 percent in this case — was more dramatic than expected (Putnam, 2003).

Replicating this study in other states would likely produce similar results. In Rhode Island, for example, the vast majority of schools needing improvement are concentrated in the urban districts of Central Falls, Newport, Pawtucket, Providence and Woonsocket. According to the *Providence Journal Bulletin*, "every urban middle school, which serves sixth through eighth graders, is in need of improvement" (Borg, 2003). In Connecticut's poorest cities, there was widespread failure; for example, all of Hartford's high schools failed to make AYP (Frahm, 2003).

Proponents of the NCLB approach to reform argue that it was precisely the intent of the law to highlight the failure of schools to serve low-income minority students. They say it is good that NCLB is shining a light on these failures because they can now be addressed.

Beyond sharing the stigma of failure and the risk of sanctions, many schools failing to meet AYP targets have very little in common with one another.

In Connecticut's poorest cities, there was widespread failure; for example, all of Hartford's high schools failed to make AYP

But what does NCLB do to address these persistent problems other than punish kids, teachers, schools and communities?

Punishing Diversity

School diversity in and of itself can be another liability under NCLB (Doyle, 2003). Economists Thomas J. Kane and Douglas O. Staiger (2001) have found that racially integrated districts will be most likely to be found wanting because of significant racial disparities in test scores. Ironically, Kane and Staiger predict, districts that have gone out of their way to integrate are likely to be sanctioned more frequently than segregated school districts. In their study, Kane and Staiger looked at states that use racial subgroup test performance to determine ratings and found that segregated schools were less likely to suffer the consequences of score variability. This is largely because the number of students in any racial group within an integrated school is likely to be so small as to make scores for the subgroup more volatile than scores for the school as a whole.

In California, for example, Kane and Staiger found that more diverse schools were substantially less likely to be rewarded for their test score gains than were more homogeneous schools, even though the more diverse schools actually had “greater improvements in overall test scores.” Thus, use of test score gains to reward or punish “can generate perverse incentives for districts to segregate their students.”

As Kane and Staiger predicted, schools *are* being punished under NCLB for being racially integrated. A report from Policy Analysis for California Education (Novak & Fuller, 2003) shows clearly that the more subgroups a school has, and the more economically disadvantaged students it enrolls, the less likely it is to make AYP.

The report shows that schools with very similar average scores fare very differently depending upon the number of subgroups they must report. In California elementary schools with 50 to 75 percent economically disadvantaged students, 71 percent of those with three subgroups made AYP, while only 55 percent of those with five subgroups did. Yet the schools with more poor kids averaged one point higher in reading and had the same average math scores. In the elementary schools serving the most low-income students, the chance of making AYP fell 30 percentage points from schools having two subgroups to those

Ironically, Kane and Staiger predict, districts that have gone out of their way to integrate are likely to be sanctioned more frequently than segregated school districts.

Schools with very similar average scores fare very differently depending upon the number of subgroups they must report.

having five subgroups (from 64 to 34 percent). Yet, on average, the latter group scored only two points lower on the state reading test and one point higher on the math exam. In short, schools performing as well as other schools are failing simply because they are more diverse.

Defenders of NCLB argue that often the reason for the difference is that while a school may be serving some students well, it may serve others less well. While this has been shown on occasion, what Kane and Staiger and the PACE report demonstrate is that much of the failure to make AYP is an artifact of NCLB's formula. For example, a low-income, limited English proficient Latino child with special needs who scores low because the child has limited English proficiency will be counted in four subgroups. Just a few such children can cause an entire school to "fail." A higher-scoring, white, English-speaking, non-poor student counts just once.

Measurement expert Robert Linn confirmed Kane and Staiger's findings that the requirement for sub-group AYP will make more schools vulnerable to being labeled "failures." Linn concludes, "The NCLB adequate yearly progress requirements represent enormous, if not overwhelming, challenges to schools, districts, and states" (Linn, 2003b).

Reports from the field demonstrate that school officials in urban districts are feeling the burden of this "diversity penalty." A study by the nonpartisan Center on Education Policy highlighted the challenge faced by urban districts with diverse, low-income student populations in a report titled "*Implementing The No Child Left Behind Act*" (Center on Education Policy, 2003). While the report in general asserted that districts are "optimistic" about their ability to meet the requirements of the law, urban districts in particular see serious obstacles to their success. "These urban systems faced special challenges in making adequate yearly progress because they tended to have more student subgroups counted for accountability purposes and more schools targeted for improvement and technical assistance. For 2003-04, Cleveland has 21 schools identified for school improvement or corrective action. To make AYP, the Cleveland public schools must show improvement every year on all 82 benchmarks in the state's AYP definition—taking into account all the subgroups, grade levels, and progress indicators counted—far more than its suburban counterparts. And because this AYP definition is based in part on state average test scores, dis-

A low-income, limited English proficient Latino child with special needs who scores low because the child has limited English proficiency will be counted in four subgroups.

"The NCLB adequate yearly progress requirements represent enormous, if not overwhelming, challenges to schools, districts, and states"

-Robert Linn

tricts with low performance, like Cleveland, must make up more ground than other districts in the state to meet the yearly benchmarks.”

Massachusetts superintendents interviewed for a report by MassINC (2003), a business-backed research group that strongly supports test-based accountability and reform, also highlighted the problem of the diversity penalty. The report’s authors write: “Superintendents noted their conviction that, because of their disproportionate impact, student subgroups with lower achievement rates are at risk of alienation, blame, and damaged self-confidence.”

Failure in the Suburbs

While few observers express surprise at seeing large numbers of high-poverty urban schools show up on lists of failing schools, the struggles of suburban schools and districts to keep up with AYP’s demands is another matter. Some NCLB proponents see suburban school failures as more proof that the law is living up to its name by identifying the pockets of poorly performing students who might otherwise go unnoticed when their scores are averaged with those of their high-performing peers. Critics question the accuracy and efficacy of tarring entire schools and districts because of the performance of specific subgroups.

Suburban school officials complain that, rather than provide support, guidance and resources to address the needs of vulnerable students in these subgroups, NCLB lays out traps and pitfalls for both excellent and neglectful schools. Some suburban districts fall victim to the diversity penalty. Many suburban schools are tripped up by the requirement that 95 percent of all students (and 95 percent of students in each subgroup) be tested. This stipulation, in particular, has led to schools that would otherwise meet and exceed expectations for improvement being labeled INOI because a handful of students were not tested. One Georgia school, for example, was labeled INOI because a single disabled student missed the state math test, meaning the school had only 42 out of its 45 special ed students take the test for a participation rate of 93, not 95 percent. If that student had taken the math exam, the participation rate would have been 96 percent and the school would have made its goals (Tofig, 2003).

“Because of their disproportionate impact, student subgroups with lower achievement rates are at risk of alienation, blame, and damaged self-confidence.”

- Massachusetts Superintendents

Rather than provide support, guidance and resources to address the needs of vulnerable students in these subgroups, NCLB lays out traps and pitfalls for both excellent and neglectful schools.

A *Contra Costa Times* computer analysis of California schools failing to make AYP found that one-third of schools and districts in Contra Costa, Alameda and Solano counties failed to make AYP because they failed to meet the 95 percent participation rate, more than for any academic reason. In the Pleasanton schools, for example, if three more English learners and two more Latino students had taken the tests, the schools would have succeeded in staying off the list of schools that had failed to make AYP (Pardington, 2003).

In Brookline, Massachusetts, an affluent yet diverse town bordering Boston, if one more Hispanic student had taken the state test, the town's one high school would have made AYP (Holland, 2003).

Conclusion

Well before its passage, those who analyzed NCLB anticipated the AYP train wreck. Now public school families are paying the price in dislocation and confusion. When schools are labeled inadequate based on the statistical idiosyncrasies of the AYP formula, the result can be the opposite of providing meaningful accountability to parents and the community. As one parent put it, "Seeing the voluminous information generated by the 'No Child Left Behind' scores left me with an unaccustomed feeling: that I had way too much information about Kentucky's schools. Unfortunately, the wealth of information provided little insight about what I, as a parent and taxpayer, am supposed to do to make schools better" (Truman, 2003).

There is a widespread consensus among researchers, educators, parents and others about the mechanics of AYP:

- Widespread failure was an inevitable outcome of its design and is being borne out in school and district results.
- There are many ways to fail under AYP, so many different kinds of schools are labeled failures.
- High-poverty schools and districts fail first, but diversity itself is punished. Failure afflicts well-off suburbs as well.
- School officials are well aware of the various ways AYP trips them up and feel frustrated rather than empowered to initiate or continue efforts toward positive change.

In Brookline, Massachusetts, if one more Hispanic student had taken the state test, the town's one high school would have made AYP

"Seeing the voluminous information generated by the 'No Child Left Behind' scores left me with an unaccustomed feeling: that I had way too much information about Kentucky's schools. Unfortunately, the wealth of information provided little insight about what I, as a parent and taxpayer, am supposed to do to make schools better"

-Kentucky Parent

B. What is “Proficiency”?

While the common perception holds that universal proficiency is a reasonable and desirable goal, it is important to understand the basis for the use of the term “proficient” in NCLB and how this could contribute to massive school failure. NCLB aims for 100 percent student proficiency by 2014. Who could object to the demand that all public schoolchildren be “proficient” in math and English? The answer depends on how you define the word. The term “proficiency” is borrowed from the National Assessment of Education Progress (NAEP) tests, which have been used for years to provide sampled snapshots of how U.S. schoolchildren are progressing in various academic subjects. NAEP’s importance will grow enormously as a result of NCLB, which requires states to use NAEP to confirm the results of their state tests.

Many newspaper reports and commentaries on NCLB and NAEP describe proficiency simply as the ability to do grade-level work. For example, a recent *Boston Globe* article on NAEP results included this explanation of what it means to score below the NAEP proficient level: “That means [students] struggle with grade-level reading and math and cannot always apply the skills to real-world situations” (Vaishnav, 2003).

A wide range of testing experts, however, have found that the definition of proficiency used by states for NCLB purposes is wildly inconsistent, and the NAEP proficiency standards are set so high that it will be impossible for most schools to reach them (Stecher *et al.*, November 2003; Linn, 2002b; Bracey, 2003). This section will address the following issues:

- The term proficiency comes from its use in NAEP testing, where it has been widely criticized for being an unrealistic and inaccurate standard.
- States vary wildly in how they define proficiency, making it difficult if not impossible to make meaningful comparisons from state to state.
- As a result, states are beginning the race to 100 percent proficiency from many different starting points, many of which do not necessarily reflect the relative academic health of their schools and students.
- Some states have resorted to lowering their standards in hopes of making the grade.

A wide range of testing experts, however, have found that the definition of proficiency used by states for NCLB purposes is wildly inconsistent, and the proficiency standards are set so high that it will be impossible for most schools to reach them

“Proficient” does not mean “grade level” or “average,” contrary to common public interpretation.

NAEP Levels

Research on the setting of NAEP “levels” (basic, proficient, advanced) shows the levels to be misleading. For example, “proficient” does not mean “grade level” or “average,” contrary to common public interpretation.

According to Professor Gerald Bracey, an independent education researcher who teaches at George Mason University, “[T]he NAEP achievement levels have been rejected by everyone who has ever studied them: UCLA’s Center for Research on Evaluation, Student Standards and Testing (CRESST), the General Accounting Office and the National Academy of Sciences, as well as by individual psychometricians such as Lyle Jones of the University of North Carolina. The studies agree that the methods used are flawed and, more importantly, the results don’t accord with any other data.

“For instance, Jones pointed out that American fourth-graders were well above average on the mathematics tests of the Third International Mathematics and Science Study (TIMSS), yet only 18 percent reached the proficient level and a meager two percent scored at the advanced level in the 1996 NAEP mathematics. Similar low percentages are seen in the 1996 NAEP Science assessment and TIMSS Science where American fourth-graders were third in the world among 26 nations. Finally, on the 2000 NAEP reading assessment, only 32 percent of fourth-graders attained proficient or better, but that [sic] American 9-year-olds were second in the world among 27 countries in the international reading study, *How in the World Do Students Read?* It makes no sense that American kids do so poorly on domestic measures such as NAEP but stack up well against the rest of the industrialized world” (Bracey, 2003).

A 1998 report from the National Academy of Sciences titled, “Grading the Nation’s Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress,” recommended that the process for setting NAEP achievement levels be replaced. “This committee, as well as the U.S. General Accounting Office, the National Academy of Education, and other evaluators, have judged the current achievement-level-setting model and results to be flawed...NAEP achievement-level results do not appear to be reasonable compared with other external information about students’ achievement” (Pellegrino *et al.*, 1998).

“American 9-year-olds were second in the world among 27 countries in the international reading study, “How in the World Do Students Read?” “ It makes no sense that American kids do so poorly on domestic measures such as NAEP but stack up well against the rest of the industrialized world”

- Gerald Bracey

NAEP achievement-level results do not appear to be reasonable compared with other external information about students’ achievement”

-National Academy of Sciences

Robert Linn and his colleagues (2002a) report that few states will meet AYP targets, precisely because the “proficient” level in most states is set too high. Perhaps more important is their contention that NCLB’s demand for steady year-to-year progress toward 100 percent proficiency is completely unrealistic. They found, for example, that only three out of 33 states posted even a one percent per year increase in the number of students scoring proficient on National Assessment of Educational Progress reading tests from 1992 to 1998. This is far below the 5 or 6 percent increase per year that would be required to reach 100 percent proficiency by 2014.

Proficiency Levels Vary Widely

State proficiency standards are not only unreasonably high, they are anything but “standard.” The Northwest Evaluation Association (NWEA) did a careful analysis of proficiency standards in 14 states and documented wild variations both among states and within states from grade level to grade level (for example, math proficiency might be set at the 46th percentile in grade 3 and at the 75th percentile in grade 8). The study generated three conclusions: there will be great variation in the percentages of students deemed proficient from state to state, even if the students have the same skills; differences in proficiency standards across grades will provide teachers with inconsistent information about students; and variations in proficiency standards between subject areas will provide inconsistent information when comparing proficiency in math and English (Kingsbury *et al.*, 2003). Variations in proficiency levels from grade to grade could have dire consequences. For example, if the level of math proficiency is set lower in grade 3 than in grade 8, a student might not be identified as needing math help until 8th grade, when it may be much harder to intervene successfully.

A Race with 50 Starting Points

A look at each state’s AYP starting points, or where states begin the race to 100 percent proficiency, gives yet another indication of the variability of state proficiency standards. Starting points are determined by using a formula based on the percentage “proficient” on state tests in 2001-02 (see Table I-2 at end of chapter). There is a startling degree of variability. Some states are faced with the challenge of reaching 100 percent proficiency in English and Language Arts by 2014 starting from lows like 13.6 percent for California high schoolers and

There will be great variation in the percentages of students deemed proficient from state to state, even if the students have the same skills.

A look at each state’s AYP starting points, or where states begin the race to 100 percent proficiency, gives yet another indication of the variability of state proficiency standards.

23 percent for high school students in Arizona. At the other end of the spectrum, Colorado students begin the race at 80.3 percent proficiency.

Are these starting points an accurate reflection of the relative academic health of these states? Professor Linn takes issue with the utility of such comparisons. In his report, “Performance Standards: Utility for Different Uses of Assessments,” Linn wrote: “State NAEP results indicate that states do vary in terms of student achievement, but not nearly enough to explain the huge variability in NCLB percentage proficient starting points. For the 43 states that participated in the 2002 NAEP 4th grade reading assessment, for example, the percentage of students who were at the proficient level or above ranged from a low of 14 percent in Mississippi to a high of 47 percent in Massachusetts” (Linn, 2003a).

The AYP gain requirements differ so markedly in part due to different state definitions of “proficient.” The NWEA study (2003) compared state-set proficiency levels in the similar, adjacent states of Wyoming and Montana. The eighth-grade math proficiency level in Montana was set at the 36th percentile, while Wyoming’s was set at the 89th percentile. This means that under NCLB more than twice as many students in Wyoming could be identified as below proficient than in Montana, even if students in both states have exactly the same achievement level on a norm-referenced test.

Shifting Standards to Make the Grade

Some states, including Texas, Arizona, Colorado and Louisiana, have lowered their definition of proficiency in hopes of increasing their chances of attaining AYP. Colorado and Louisiana created a dual system, with a lesser definition of “proficient” for meeting federal requirements and a more stringent definition for local use (FairTest Examiner, Fall 2002). Rod Paige, U.S. Secretary of Education, denounced these moves in a letter to state education superintendents, yet Colorado was one of the first states to have its NCLB plans approved by his agency, the U.S. Department of Education.

In other cases, state accountability requirements are easier to meet than AYP requirements. For example, in Arizona, 289 schools were identified as needing improvement under NCLB, but these same schools met the state’s performance targets and earned either a “performing” or “highly performing” label. In Virginia, 723 (40 percent of

Under NCLB more than twice as many students in Wyoming could be identified as below proficient than in Montana, even if students in both states have exactly the same achievement level on a norm-referenced test.

Some states, including Texas, Arizona, Colorado and Louisiana, have lowered their definition of proficiency in hopes of increasing their chances of attaining AYP.

all schools) failed to make federal AYP goals while only 402 (22 percent) failed to meet state accreditation standards (Sunderman and Kim, 2004). (See Appendix 1: “Jeb’s A+ Schools Dubbed Failing by George,” and Appendix 2: “When AYP Means Good Schools Are Called Bad” at the end of this chapter.)

Such dual standards are likely to increase confusion among parents about the meaning of the various ratings. Dual systems could also increase administration costs if larger staffs will be needed to compile and analyze two sets of statistics.

Some state education policymakers could be holding onto the original state system because they are betting NCLB will be overhauled before 2014. “If you chuck your entire state system and later on the federal law does change, then you’re left with an unworkable system,” said William Padia, director of policy and evaluation at the California Department of Education. “Better to live with an uncomfortable marriage of the two” (Galehouse, 2003).

In response to the problems implementing NCLB during its first two years and after seeing that different states have widely varying compliance agreements with the federal government, more than 40 states have requested permission from the U.S. DOE to make changes in their NCLB accountability plans (Olson, 2004). Connecticut wants to test only in grades 4, 6, 8, and 10, rather than 3-8 inclusive. Other states plan to revise the minimum size of the “cells” used to determine whether a subgroup will be counted in AYP determinations. A number of states want to use “growth” models (student progress) rather than the absolute score requirements set through NCLB. Such approaches could allow schools to gain credit for students’ partial progress toward the proficient level, slowing down the speed at which they approach “in need of improvement” status. Some states want to be able to count students who fit multiple subpopulations (e.g., Latino, limited English proficient, low-income and disabled) in just one subgroup for AYP calculations.

Problems with the definition of “proficiency” mean that NCLB’s goal of bringing all students up to “proficiency” is anything but reasonable.

Such dual standards are likely to increase confusion among parents about the meaning of the various ratings.

Some state education policymakers could be holding onto the original state system because they are betting NCLB will be overhauled before 2014.

- By borrowing from NAEP terminology and using the NAEP levels to double-check state progress, the nation begins with a false assumption about how well the majority of U.S. students are doing.
- The variability in defining proficiency means that schools, educators, students and their families face severe consequences in some states that they would not face in other states with similar school quality and student attainment.
- Some states are responding to the unreasonable demands of NCLB by lowering their standards in hopes of avoiding sanctions.

Errors are common in standardized testing and are likely to skyrocket with the explosion of testing brought about by NCLB.

C. What's Wrong with Reliance on Test Scores?

A key reason NCLB sets schools up to fail is its almost total reliance on standardized test scores to judge the success or failure of public schools. What's wrong with using test scores alone? A number of independent researchers have investigated this question and reached the following conclusions:

- Test score fluctuations are often a matter of luck rather than real progress.
- A certain level of failure is built into the design of most standardized tests.
- Errors are common in standardized testing and are likely to skyrocket with the explosion of testing brought about by NCLB.

(The next chapter will analyze the dangers to curriculum and instruction caused by overreliance on standardized test scores.)

Score Swings a Matter of Luck

School officials and politicians trumpet each bump upward in test scores and respond to such movement with rewards and still more test-driven education policy. Reaction tends to be more muted when the scores that have gone up eventually come down again. Researchers Kane and Staiger (2001a, 2001b) warned about these problems even before NCLB passed Congress. Their research into the reasons for test score fluctuations, for example, consistently found that year-to-year gains and losses on state tests are too unreliable to be used for decision-making.

Year-to-year gains and losses on state tests are too unreliable to be used for decision-making.

In two papers Kane and Staiger examined data, primarily from North Carolina, to determine the precision of school scores. They categorized test score variations as due to sampling changes (e.g., a different group of students each year in a tested grade), a particularly severe problem in small schools; “one-time factors” such as a barking dog that distracts a group of test-takers; and persistent differences in actual performance among schools.

The researchers found that 50 to 80 percent of the year-to-year observed fluctuation in a typical North Carolina school’s average score is due to the first two factors, not differences in tested achievement.

As a result, school rankings based largely on score increases “generally resemble a lottery.” Only one percent of the state’s schools ranked in the top 10 percent in math for all six of the years studied. In reading, which is more volatile than math, more than one-third of all schools ranked in the top 10 percent at some point.

Selecting “good” programs that other schools should emulate, a common goal of test-based accountability programs, is also a matter of luck. If test scores are the determining factor, a large percentage of schools will at some point be labeled “best practice” schools. The result would be an ever-expanding menu of “best practices” from schools whose scores often decline over the next year or two.

Another researcher, Boston College education professor Walter Haney, looked at all Massachusetts elementary schools that showed a 10-point gain on the state test from 1999-2000 and found that most posted score declines in 2001, often as large as the gains of the previous year. The data showed that a school that did better the first time was more likely than not to do worse the second time.

According to Haney, “These results don’t mean that teachers or students became lazy and tried to coast on their success. They mean that there was never really evidence of success at all. Particularly in small schools, as other research has confirmed, changes in score averages from year to year are poor measures of school quality. If fewer than 100 students are tested in each grade, averages may swing widely from year to year simply because of the particular samples of students tested and the vagaries of annual test content and administration” (Haney, 2002).

School rankings based largely on score increases “generally resemble a lottery.”
- Kane and Staiger

Selecting “good” programs that other schools should emulate, a common goal of test-based accountability programs, is also a matter of luck.

If too few students are in a school or subgroup that must be measured for its progress, the results will be particularly volatile. However, most states use minimum subgroup sizes that are far smaller than those that Haney and other testing experts say must be used to attain statistically sound results. Among the states whose AYP plans were initially approved by the U.S. Department of Education, Colorado, Indiana and Ohio will use 30 as a minimum group size. Massachusetts will use only 20. (See Table I-3 and Chart I-1 at end of chapter.)

If subgroup sizes are too small, some schools will fall short of AYP due to measurement error, not because of any academic problem. Some states will use statistical procedures to ameliorate the effects of too-small groups. These states determine the margin of error for each subgroup, based on the number of students tested and the percent who reach proficiency, and use it to give themselves more leeway in determining how many schools have failed to make AYP. Kentucky is one state using a so-called confidence-interval formula. A *Lexington Herald-Leader* reporter described how such a process works: “If a Kentucky school had 17 African-American students and three of them — 18 percent — reached proficiency, the margin of error for that subgroup would be plus or minus 28 percentage points. Those points would be added to the raw score, and a school would be evaluated as if 46 percent of students had scored proficiently” (Deffendall, 2003). While this makes statistical sense and helps prevent over-identification of schools as failing, it typically mystifies the public, which often concludes states are manipulating numbers to hide failure.

In the hopes of making more accurate judgments, some states have taken to considering multi-year, moving averages rather than just the previous year’s scores. However, as Kane and Staiger report (2001a, 2001b), a North Carolina school which desired to predict its current year reading score gains would be better off to simply pick the state’s average score increase rather than to use its own previous four years of score changes. Thus, averaging a few years’ scores in an effort to solve the problem of random fluctuations appears not to sufficiently reduce misinformation.

The authors also asked whether score gains are due to such factors as teaching to the test rather than to real improvements in learning. Lacking a direct measure, they examined several characteristics of “student engagement” -- absenteeism, time doing homework, and time watching television. Those measures did not improve in schools in

If subgroup sizes are too small, some schools will fall short of AYP due to measurement error, not because of any academic problem.

Schools began tailoring their curricula to improve performance on the tests, without generating similar improvements on other measures.

which scores rose substantially. The authors note that this lack of improvement “would be consistent with the hypothesis that schools began tailoring their curricula to improve performance on the tests, without generating similar improvements on other measures.”

Failure Built In to Test Design

Some state tests are constructed to guarantee that a certain percentage of students will fail every year. This is because the test design methodology commonly employed in state exams rests on a technology used to develop norm-referenced tests, which sort and rank test-takers, always leaving some students at the bottom of the curve. Low-scoring students are disproportionately low-income, African-American, Latino, recent immigrants whose first language is not English, and students with disabilities.

In “Ensuring Failure,” Haney (2002) wrote: “When questions answered correctly by more than 70 percent of students are systematically excluded from the exam, this guarantees continuing failure. Tests like the MCAS [the Massachusetts Comprehensive Assessment System exams] are designed so that all students can never succeed. Evidence suggests that other state tests (in Texas, California, and New York, for example) also have been constructed using norm-referenced test-construction procedures.” In fact, research conducted for a lawsuit challenging the Texas graduation test found the TAAS exam had the flaws described by Haney (FairTest *Examiner*, 1999-2000).

Error-Prone Tests

The New York Times (Henriques, 2003) reported in detail on the prevalence of errors in standardized testing. A series of articles warned that the sharp increase in testing volume created by NCLB may cause a spike in human errors unless greater attention is paid to quality control issues.

Prof. Mark L. Davison, an educational psychologist at the University of Minnesota, predicted a doubling of testing in the next few years as a result of NCLB. “I think preventing [errors] entirely is impossible,” he told the *Times*. “As existing companies expand and new companies move into the field, they’re going to experience growing pains.”

Tests like the MCAS [the Massachusetts Comprehensive Assessment System exams] are designed so that all students can never succeed.

-Walt Haney

The sharp increase in testing volume created by NCLB may cause a spike in human errors unless greater attention is paid to quality control issues.

A report by the National Board on Educational Testing and Policy based at Boston College, *Errors in Standardized Tests: A Systemic Problem*, highlights the nature and extent of human mistakes in educational testing over the past 25 years (Rhoades and Madaus, 2003). In contrast to random measurement error expected in all tests, human error is unexpected and brings unknown, often harmful consequences for students and schools, including:

- Inaccurately preventing high school seniors from receiving a diploma (Minnesota 2000);
- Creating misleading “worst school” lists (Pennsylvania 1996, Nevada 1999, Ohio 2002);
- Erroneously assigning students to remedial classes or retaining them in grade (New Jersey 1993, New York City 1999, Maryland, 2001);
- Barring qualified college applicants from attending their chosen universities (Scotland 2000, England 2002); and
- Denying qualified applicants access to professional credentials (Alabama 1981-85, New York 1981, Oklahoma 2000).

Rhoades and Madaus point out that these errors occur in an industry whose activities are largely unregulated, an environment where mistakes are difficult to detect. As the amount of testing has increased, the industry has been spread thin and testing errors have risen. NCLB’s mandated increase in testing is likely to cause a larger jump in the number of errors in designing tests, setting passing scores, establishing norm groups, scoring exams, and reporting results.

The Boston College report demonstrates that testing is a fallible technology, subject to internal and external errors. With errors an unavoidable problem, basing important educational decisions on the outcome of one test can put children and schools at risk due to foul-ups that may never be caught or remedied.

The release of Illinois’s NCLB report card — with an estimated 34,261 mistakes involving about 75 percent of the state’s schools, according to the *Chicago Tribune* — offers a vivid illustration of the Boston College report’s conclusions. The Tribune said there were so many errors, “it is virtually impossible to draw meaningful conclusions that educators had hoped for and that the federal No Child Left Behind Act requires.” State school officials said 368 schools may have been mistakenly placed on a federal failure list because of data errors.

These errors occur in an industry whose activities are largely unregulated, an environment where mistakes are difficult to detect.

There were so many errors, “it is virtually impossible to draw meaningful conclusions that educators had hoped for and that the federal No Child Left Behind Act requires.”

-Chicago Tribune

The list reportedly included several well-known, high-performing elementary and high schools (Banchero and Little, 2003)

Summary

For President Bush and other NCLB proponents, the law's near-total reliance on test scores to determine the progress of students, teachers and schools reflects a desire for "objective" assessments of how schools are doing. Bush says, for example, "Without yearly testing, we don't know who is falling behind and who needs help. Without yearly testing, too often we don't find failure until it is too late to fix" (Bush, 2001). But standardized test scores, while they have some utility, offer nothing more than a snapshot of student achievement at a moment in time and can be misleading when used to make important decisions about students and schools. The national focus, some would say obsession, on standardized test scores to drive school improvement and reform is not an entirely new policy. The historic record casts serious doubt on the decision to continue a nationwide experiment in test-based reform. Among the findings:

- Test score fluctuations do not necessarily indicate real progress when scores rise or deterioration when they fall and should not be used by themselves to reward or sanction schools, teachers or school officials.
- Many of the tests that are being used to judge our students, teachers and schools are specifically designed to ensure a certain proportion of failures.
- Errors have always been a part of standardized testing and are likely to increase substantially with the increase in testing mandated by NCLB.

D. Conclusion

Most NCLB critics do not deny that there are public schools offering inadequate educations to their students. On the contrary, they say, there are still too many schools that, for a variety of reasons, are not giving students what they need to succeed. NCLB, however, will exacerbate this situation by labeling so many schools as underperforming that it will be impossible to really identify schools that need improvement and give them what they need to improve. As the National Education Association's Joel Packer put it, "It's like saying everybody in the country is sick. How do you figure out how to focus

"It's like saying everybody in the country is sick. How do you figure out how to focus your resources, especially if what you want to do is help those who are really sick?"

-National
Education
Association

"The adequate yearly progress [AYP] net has been cast very wide, and so it's going to catch a very high percentage of schools"

your resources, especially if what you want to do is help those who are really sick?” (National Education Association, 2003).

David Shreve, an education policy expert at the National Conference of State Legislatures, concurs with this view (Marks, July 21). “The adequate yearly progress [AYP] net has been cast very wide, and so it’s going to catch a very high percentage of schools,” says Shreve. “We either have to accept the fact that a vast majority of our schools are awful, or we have to accept the AYP net has been cast too broadly and it’s catching way more schools than it should.”

While the AYP formula is mind-bogglingly complex and will have a range of complicated consequences for public schools, behind it is a simplistic view that most educators are not working hard and that improvement will result when they feel the “tough love” of a kick in the pants.

Most of those who work in and attend these schools see a vastly different reality: many challenges and dwindling resources. The people who are charged with reaching the 100 percent proficient goal see a law that is heavy on punishment and light on the tools they need to do the job. “There’s no guarantee whatsoever under No Child Left Behind that any school has the basic resources that they need to bring these children up to the level of achievement the law calls for,” says Michael Rebell, executive director and counsel of the Campaign for Fiscal Equity, a nonprofit education funding advocacy group in New York. “You can beat them over the head as much as you want, but you can’t get blood from a stone” (Marks, 2003).

Or, as the principal of a predominantly black North Carolina charter school put it: “I’m not so sure this law is about leaving no child behind.” Jackie M’Buru, principal of SPARC Academy in Raleigh, added, “I think this law is about blaming teachers and principals who need more support. If you say you need more help, the answer is: ‘Get better or we’ll shut you down’” (Simmons, 2003).

“There’s no guarantee whatsoever under No Child Left Behind that any school has the basic resources that they need to bring these children up to the level of achievement the law calls for,”

-Michael Rebell

“I think this law is about blaming teachers and principals who need more support. If you say you need more help, the answer is: ‘Get better or we’ll shut you down’”

-Jackie M’Buru

Appendix 1: Jeb's A+ Schools Dubbed Failing by George

If anyone was going to get President George W. Bush's NCLB test right, you'd think it would be his brother, Florida Governor Jeb Bush. So are Florida's schools models of the improvements to be reaped by test-driven reform? That depends on which Bush brother you ask (Pinzur, 2003).

Florida Governor Jeb Bush touted improvement based on Florida's ranking system, with six times as many A schools in 2003 than when grading began in 1999 and fewer than half as many F schools. But within months, President George Bush's NCLB report card found that of 1,229 A schools statewide, only 22 percent made Adequate Yearly Progress.

Six percent of the state's B schools made AYP, as did two percent of C schools. No D or F schools qualified. Florida education officials scrambled to explain. 'Just like an `A' student has room for improvement, even a top school can work toward improving performance,' said Frances Marine, spokeswoman for the Florida Department of Education. But the contradictory picture produced by the state and federal ranking systems was a perfect illustration of how unenlightening it is when accountability is boiled down to standardized test results. Parents are left struggling to make sense of it all.

Education leaders are especially concerned about confusing parents, who have heard the governor celebrate annual improvements in school grades. 'If I saw that my state graded me as an `A' and then the federal government said we hadn't met the No Child Left Behind Act, I would be very confused and asking a lot of questions,' said Karin Brown, a parent activist and former president of the Dade County Council PTA/PTSA. 'From a parent point of view, there's definitely a contradiction here.'

An editorial writer for the Bradenton Herald summed it up: 'The answer, of course, is that neither represents a fair and accurate picture of the quality of education being provided to students here or elsewhere in the state, which had a 90 percent failure rate in the federal test. Rather, they are snapshots of student performance as measured by an arbitrary set of standards. But they don't necessarily reflect teachers' success at educating children with widely varied levels of ability, socialization and language mastery. Throw in a different set of standards and you likely would get yet another, entirely different result' (Editorial, 2003).

The contradictory picture produced by the state and federal ranking systems was a perfect illustration of how unenlightening it is when accountability is boiled down to standardized test results.

Appendix 2: When AYP Means Good Schools Are Called Bad

The rigidity, complexity and insatiable demands of NCLB's AYP formula mean that schools across the country that had been lauded for improvement and excellence have unceremoniously landed on lists of schools "In Need of Improvement."

One was Southfield, Michigan's Vandenberg Elementary, which President Bush had visited to promote ESEA. *USA Today* found that 19 U.S. Department of Education Blue Ribbon exemplary schools ended up on low-performing lists (Thomas and DeBarros, 2003).

Mark Christie, the Republican former president of the Virginia Board of Education, has decried the way NCLB labels many excellent schools as INOI. "If you create a system that calls a good school a bad school, people will know and lose faith in accountability," Mr. Christie said.

Christie was quoted in a New York Times article by Michael Winerip (2003) that highlighted the experience of one Virginia school, Tuckahoe Middle School in suburban Henrico County. According to Winerip, "Tuckahoe's test scores are among the best in Virginia, with 99 percent achieving proficiency in math, 95 percent in English. Its previous principal was the 2002 state principal of the year, and in 1996 Tuckahoe was named a Blue Ribbon School of Excellence by the federal Education Department." How did such a school end up on the list of schools INOI? It missed by one percentage point the mandate that 95 percent of students be tested because it had recent Bosnian immigrants who didn't speak English well enough to be tested in English and the state did not have a test to give them as an alternative. "It didn't make sense to have them take a test they couldn't understand," said Kurt Hulett, Tuckahoe's principal.

Of course, Tuckahoe is far from an isolated example. In Tennessee, about 160 schools received incentive grants of \$5,150 last year for meeting state benchmarks. A few months later, 40 of those schools wound up on this list of schools that failed to make AYP (Riley, 2003).

Kentucky found more than a quarter of its schools, 470, failed to make AYP last year. Forty of those same schools had been recognized

USA Today found that 19 U.S. Department of Education Blue Ribbon exemplary schools ended up on low-performing lists

Kentucky found more than a quarter of its schools, 470, failed to make AYP last year. Forty of those same schools had been recognized just one month earlier for being a year ahead of their academic goals.

just one month earlier for being a year ahead of their academic goals set by the Kentucky state testing system. Emma B. Ward Elementary School was one of the schools recognized for improvement by the state and called failing by the feds. School officials complained that the federal designation was out of date and did not reflect the improvements they had already made. Principal Sarah Sweat said, “It’s really unfortunate. We know that what we did last year got us to the score we got, which was the highest in the district. We know we are doing the right thing to achieve our goals” (Rodriguez, 2003).

The feds have come up with a solution to the embarrassing problem of Blue Ribbon Schools being labeled failures by NCLB. While in the past, Blue Ribbon Schools were evaluated on multiple measures, including school visits and interviews with support staff, now getting a Blue Ribbon will depend solely on test scores.

While in the past, Blue Ribbon Schools were evaluated on multiple measures, including school visits and interviews with support staff, now getting a Blue Ribbon will depend solely on test scores.

References

- Banchero, S. and Little, D. December 19, 2003. "Errors Fill State Testing Data," *Chicago Tribune*.
- Borg, L. October 10, 2003. "Schools Decline in Annual Rankings," *Providence Journal Bulletin*.
- Bracey, G. February 2003. "The No Child Left Behind Act: Just Say No." Available online at <http://www.nochildleft.com>.
- Bradenton (FL) Herald*. August 10, 2003. "Accountability Sham: State, Federal Tests Results Conflict."
- Bush, G. W. January 23, 2001. Press Conference with President Bush and Education Secretary Rod Paige to Introduce the President's Education Program. Available on the web at <http://www.whitehouse.gov/news/releases/2001/01/20010123-2.html>.
- Center on Education Policy. October 2003. "Implementing The No Child Left Behind Act." Available online at www.cep-dc.org.
- Center on Education Policy. January 2004. "From the Capital to the Classroom." Available online at www.cep-dc.org.
- Deffendall, L. October 19, 2003. "No Statistics Are Being Left Behind," *Lexington Herald-Leader*.
- Dillon, S. January 2, 2004. "Some School Districts Challenge Bush's Signature Education Law," *The New York Times*.
- Doyle, D. P. April 16, 2003. The Doyle Report, Issue 3.16, No. 63. Available online at [http://www.thedoylereport.com/cyber_chair?object=archive\[\]&content_id=3696](http://www.thedoylereport.com/cyber_chair?object=archive[]&content_id=3696).
- Editorial. August 10, 2003. "Accountability Sham: State, Federal Test Results Conflict," *Bradenton Herald (FL)*.
- Elmore, R. F. November 2003. "A Plea for Strong Practice," *Education Leadership*, Volume 61, Number 3.
- eSchool News. October 1, 2003. "A Look at School AYP Failure by State." Available online at <http://www.eschoolnews.com/news/issue.cfm?PubID=1&IssueID=201>.
- FairTest *Examiner*. Winter 1999-2000. "Court Rules for High-Stakes Testing." Available online at <http://www.fairtest.org>.
- FairTest *Examiner*. Summer 2002. "ESEA: Ten Percent of U.S. Schools Labeled 'Failing.'" Available online at <http://www.fairtest.org>.
- Frahm, R. A. December 18, 2003. "46% of Schools 'Left Behind'," *Hartford Courant*.
- Fuetsch, M. August 12, 2003. "Delaware Schools Reeling: Unforgiving Federal Ratings Fail to Show Actual Progress," *The News Journal*.
- Galehouse, M. July 14, 2003. "Academic Bar Lowered to Get Schools on Track," *The Arizona Republic*.

- Gootman, E. September 11, 2003. "40 Percent of City Schools Do Not Meet U.S. Standards," *The New York Times*.
- Haney, W. July 10, 2002. "Ensuring Failure," *Education Week*.
- Helfand, D. January 9, 2003. "State Keeps Education Standards," *Los Angeles Times*.
- Henriques, D. B. September 2, 2003. "Rising Demands for Testing Push Limits of Its Accuracy," *The New York Times*.
- Holland, R. December 10, 2003. "Some Schools Miss MCAS Progress Targets," *Brookline Tab*.
- Howard, S. December 1, 2003. Quoted in Susan Ohanian, "Bush Flunks Schools," *Nation*.
- Hutton, T. July 2003. "Critics Argue that NCLB's Adequate Yearly Progress Provisions Mislabeled Schools," National School Boards Association. Available online at http://www.nsba.org/site/doc_cosa.asp?TRACKID=&VID=50&CID=1046&DID=31716
- Jacques, N. January 31, 2004. "The Grades Are In," *Battle Creek Enquirer*.
- Kane, T. J. and Staiger, D. O. April 2001. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." Paper presented at a Brookings Institution Conference.
- Kane, T. J. and Staiger, D. O. March 2001. "Improving School Accountability Measures." Available online at <http://papers.nber.org/papers/W8156>.
- Kingsbury, G. et al. November 24, 2003. "The State of Standards," Northwest Evaluation Association. Available online at <http://www.young-roehr.com/nwea/>.
- Lee, J. April 7, 2004. "How Feasible is Adequate Yearly Progress (AYP)? Simulations of School AYP 'Uniform Averaging' and 'Safe Harbor' under the No Child Left Behind Act." *Education Policy Analysis Archives*, V. 12, N. 14. Available online at <http://epaa.asu.edu/epaa/v12n14/>.
- Linn, R. L., Baker, E., Herman, J. Spring 2002. *The CRESST Line*. Available online at <http://cresst96.cse.ucla.edu/products/newsletters/CLSpring02final.pdf>.
- Linn, R. L., Baker, E. L., and Betebenner, D. W. August/September 2002. "Accountability Systems: Implications of the No Child Left Behind Act of 2001." *Educational Researcher*.
- Linn et al. Fall 2002. "Minimum Group Size for Measuring Adequate Yearly Progress," *The CRESST Line*. Available online at <http://www.cse.ucla.edu/products/newsletters/CL2002fall.pdf>.
- Linn, R. L. September 1, 2003. "Performance Standards: Utility for Different Uses of Assessments," *Education Policy Analysis*

- Archives*, 11(31). Available online at <http://epaa.asu.edu/epaa/v11n31/>.
- Linn, R. L. Winter 2003. "Requirements for Measuring Adequate Yearly Progress," Policy Brief 6, National Center for Research on Evaluation, Standards, and Student Testing.
- MacDonald, C. November 30, 2003. "U.S. Flunks Top Metro Schools," *The Detroit News*.
- Maine Education Association. March 2003. "ESEA Stinks," News and Views, *Online MEA*. Available online at http://www.maine.nea.org/dir2/esea_stinks.htm.
- Marks, A. July 21, 2003. "As Schools 'Fail,' Parents Talk Transfers," *The Christian Science Monitor*.
- Massachusetts Department of Education web site. 2003. Available online at <http://www.doe.mass.edu/news/news.asp?id=1714>.
- MassINC. Fall 2003. The Center for Education Research and Policy, "No Child Left Behind: Opportunity or Obstacle for Massachusetts?" Available online at www.massinc.org/about/serp/research/NCLB%20Superintendent%20Survey.pdf.
- Mathews, J. January 27, 2004. "What the Media Are Missing: Reports of Average Test Scores Mask Improvements Made by Minorities," *The Washington Post*.
- Miller, K. September 10, 2003. "U.S. Education Law Nets Mishmash of Results," *Palm Beach Post*.
- Montoya, B. October 10, 2003. "No Child Left Behind: Schools Fall Short of Goals," *Las Vegas Review-Journal*.
- Pellegrino, J. W., Jones, L. R., and Mitchell, K.J. (Eds.). 1998. National Academy of Sciences, *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Washington, DC: National Academy Press.
- National Education Association. May 2003. "No Child Left Behind?" *NEA Today*.
- Novak, J. R. and Fuller, B. December 2003. "Penalizing Diverse Schools? Similar Test Scores but Different Students Bring Federal Sanctions," *Policy Analysis for California Education (PACE)*, Policy Brief 03-4. Available online at http://pace.berkeley.edu/pace_publications.html.
- Olson, L. May 5, 2004. "States Seek Federal OK for Revisions," *Education Week*.
- Pardington, S. November 9, 2003. "Education Overhaul's Unintended Outcome," *Contra Costa Times*. Available online at <http://www.bayarea.com/mld/cctimes/living/education/7220820.htm>

- Pinzur, M. I. August 8, 2003. "State Schools Fail to Meet New Federal Test Standards," *Miami Herald*.
- Prah, P. December 9, 2002. "New Rules May Guarantee 'F's For Many Schools," *Stateline.org*.
- Prah, P. June 16, 2003. "States Get Leeway to Meet Education Law," *Stateline.org*.
- Putnam, J. July 11, 2003. "Nearly All Troubled Schools in High-Poverty, Minority Neighborhoods," *Ann Arbor News*.
- Rhoades, K. and Madaus, G. May 2003. "Errors in Standardized Tests: A Systemic Problem," National Board on Educational Testing and Public Policy. Available online at <http://www.bc.edu/nbetpp>.
- Riley, C. November 16, 2003. "Academic Reward System to Spotlight State's Top Schools," *The Tennessean*.
- Rodriguez, N. C. November 18, 2003. "470 Schools Fall Short on New Standards," *The Courier-Journal*.
- School Funding Services. January 30, 2004. Available online at http://www.schoolfundingservices.org/news/hot_tips.01.30.04.pdf.
- Simmons, T. July 30, 2003. "Law Threatens Some Schools," *The News & Observer*.
- Stecher, B. *et al.* 2003. "Working Smarter to Leave No Child Behind: Practical Insights for School Leaders," RAND Corp.
- Sunderman, G. L. and Kim, J. February 2004. "Large Mandates and Limited Resources: State Response to the No Child Left Behind Act and Implications for Accountability," Harvard Civil Rights Project.
- Thomas, K. and DeBarros, A. August 5, 2002. "School 'Excellence' Thrown a Grading Curve," *USA Today*.
- Tofig, D. August 7, 2003. "'Listed' Schools: Some See Stigma in 3Numbers Game," *Atlanta Journal Constitution*.
- Truman, C. November 19, 2003. "What the School Scores Really Tell," *Lexington Herald Leader*.
- Vaishnav, A. December 5, 2003. "School Ratings Raise Concern for Hispanics," *The Boston Globe*.
- Winerip, M. October 8, 2003. "How a Good School Can Fail on Paper," *The New York Times*.

Table 1-1: Data on Schools Not Making Adequate Yearly Progress (AYP)

Source: State Department of Education websites and press reports. States also periodically revise their lists. Because of variations in state terminology and data reporting, these numbers are not completely comparable across states.

STATE	# OF PUBLIC SCHOOLS (used to make AYP decisions)	# OF SCHOOLS NOT MAKING AYP AT LEAST ONE YEAR	% OF ALL PUBLIC SCHOOLS	# OF SCHOOLS IN SCHOOL IMPROVEMENT, CORRECTIVE ACTION, OR RESTRUCTURING (Not making AYP 2 or more years)	% OF ALL PUBLIC SCHOOLS	Updated
AL	1,547	71	4.6%	46**	2.97%	2/17
AK	488	282	57.8%	66	13.5%	2/4
AZ	1,695	403	23.8%	219	12.92%	2/17
AR	1,130	281	24.9%	17	1.5%	3/25
CA	8,710	3,947	45.3%	1,991	22.86%	2/24
CO	1,613	436	27.0%	82	5.08%	2/17
CT	1,079	230	21.3%	14	1.3%	3/23
DC	151	83	55.0%	15	10.1%	4/5
DE	171	97	56.7%	12	7.0%	3/9
FL	3,182	2,466	77.5%	45	1.41%	2/17
GA	1,999	730	36.5%	257	12.9%	
HI	280	168	60.0%	84	30.0%	4/9
ID	645	428	66.4%	43	6.67%	2/17
IL	3,919	1,688	43.1%	577	14.72%	2/17
IN	1,891	97*	5.13%	50*	2.64%	
IA	1,442	156	10.8%	11	0.76%	2/17
KS	1,413	175	12.4%	30	2.1%	3/3
KY	1,179	470	39.9%	25	2.12%	2/17
LA	1,375	620	45.1%	69	5.0%	3/25
ME	707	167	23.6%	10	1.41%	4/13
MD	1,403	518	36.9%	131	9.3%	3/5
MA	1,694	625	36.9%	208	12.3%	4/13
MI	3,472	896	25.8%	363	10.5%	3/5
MN	1,949	155	8.0%	38	2.0%	3/31
MS	870	220	25.3%	7	0.8%	3/5
MO	2,055	1,009	49.1%	32	1.56%	4/15
MT	858	212	24.7%	40	4.7%	3/5
NE	1,220	269	22.1%	6	0.5%	4/9
NV	517	221	42.8%	21	4.1%	4/7
NH	448	140	32.1%	11	2.5%	3/1
NJ	2,448	1,051	42.9%	264	10.8%	

Table 1-1: Continued

STATE	# OF PUBLIC SCHOOLS (used to make AYP decisions)	# OF SCHOOLS NOT MAKING AYP AT LEAST ONE YEAR	% OF ALL PUBLIC SCHOOLS	# OF SCHOOLS IN SCHOOL IMPROVEMENT, CORRECTIVE ACTION, OR RESTRUCTURING (Not making AYP 2 or more years)	% OF ALL PUBLIC SCHOOLS	Updated
NM	780	164	22.0%	95	12.2%	3/25
NY	4,186	1,047	25.0%	729	17.4%	4/5
NC	2,252	1,194	53.0%	36	1.6%	3/25
ND	497	47	9.5%	23	4.6%	3/9
OH	3,715	815	21.9%	191	5.1%	3/29
OK	1,796	405	22.3%	46	2.6%	4/13
OR	1,231	333	27.1%	7	0.57%	3/9
PA	2,768	1,091	39.4%	298	10.8%	3/25
RI	313	98	31.31%	27	8.6%	3/25
SC	1,072	817	76.21%	90	8.4%	3/31
SD	737	196	26.6%	32	4.3%	4/12
TN	1,650	746	45.2%	62	3.8%	4/15
TX	7,733	563	7.3%	9	0.12%	3/9
UT	865	246	28.4%	6	0.7%	3/9
VT	307	39	12.7%	9	2.9%	3/9
VA	1,806	740	41.0%	66	3.7%	3/25
WA	1,613	436	27.3%	50	3.1%	3/15
WV	728	295	40.5%	33	4.5%	3/11
WI	2,019	171	8.5%	68	3.4%	3/11
WY	364	55	15.1%	0	0%	3/8
US TOTAL	87,982	27,712	31.50%	6,565	7.46%	

* State only released list of schools not making AYP for two or more years, and did not release list of schools not making AYP for just one year.

** Title I schools only.

Prepared by National Education Association Great Public Schools Action Plan. Used with permission of NEA.

Table 1-2: State AYP Starting Points

State		Reading/ELA	Mathematics	State		Reading/ELA	Mathematics
Alaska		64.03	54.86	Nebraska	Gr 4	62	65
Arizona	Gr 3	44	32		Gr 8	61	58
	Gr 5	32	20		Gr 11	66	62
	Gr 8	31	7	Nevada	Elementary	32.4	37.3
	High	23	10		Middle	37	38
Arkansas	K-5	31.8	28.2		High	91	58
	Gr 6-8	18.1	15.3	New Hampshire	Gr 3-8	60	70
	Gr 9-12	19	10.4		High	64	52
California	Gr 2-8	13.6	16	New Jersey	Gr 4	68	53
	High	11.2	9.6		Gr 8	58	39
Colorado	Elementary	77.5	79.5		Gr 11	73	55
	Middle	74.6	60.7	New York*	Elementary	123	136
	High	80.3	50.5		Middle	107	81
Connecticut	Elementary	55	64		High	142	132
	Middle	55	64	North Carolina	Gr 3-8	68.9	74.6
	High	62	59		Gr 10	52	54.9
DC	Elementary	30.3	38.4	North Dakota	Gr 4	65.1	45.7
	High	13.7	19.8		Gr 8	61.4	33.3
Delaware		53.9	30		Gr 12	42.9	24.1
Florida		30.68	37.54	Ohio	Elementary	40.5	35.9
Georgia	Gr 3-8	60	50		Middle	36	36.8
	High	88	81		High	78	53.1
Hawaii		30	10	Oklahoma		62.2	64.8
Idaho		66	51	Oregon		40	39
Illinois		40	40	Pennsylvania		45	35
Indiana		59	57	Rhode Island	Elementary	76.1	61.7
Iowa	Elementary	65	64		Middle	68	46.1
	Middle	61	63		High	62.6	44.8
	High	69	69	South Carolina		17.6	15.5
Kansas	K-8	51.2	46.8	South Dakota	Elementary	65	45
	High	44	29.1		High	50	60
Kentucky	Elementary	47.5	22.73	Tennessee	Elementary	77.1	72.4
	Middle	45.6	16.51		High	86	65.4
	High	19.26	19.84	Texas		46.8	33.4
Louisiana		36.9	30.1	Utah	Gr 3-8	65	57
Maine	Elementary	34	12	Vermont*	Grade Span		Start Pt.
	Middle	35	13		Math	2, 4	314
	High	44	11		ELA	2, 4	385
Maryland	Gr 3	58.1	65		Math	8	287
	Gr 5	65.7	55		ELA	8	342
	Gr 8	59.9	39.7		Math	2, 4, 8, 10	293
	High	61.4	43.4		ELA	2, 4, 8, 10	380
Massachusetts		39.7	19.5		Math	8, 10	277
Michigan	Elementary	38	47		ELA	8, 10	339
	Middle	31	31		Math	10	268
	High	42	33		ELA	10	345
Minnesota	Gr 3	62.75	66.17		Math	2, 4, 8	306
	Gr 5	69.89	65.35		ELA	2, 4, 8	381
	High			Virginia		60.7	58.4
Mississippi	Gr 3	61	72	Washington	Gr 4	52.2	29.7
	Gr 4	66	49		Gr 7	30.1	17.3
	Gr 5	58	35		Gr 10	48.6	24.8
	Gr 6	51	39	Wisconsin		61	37
	Gr 7	36	19	Wyoming	Gr 4	30.4	23.8
	Gr 8	30	23		Gr 8	34.5	25.3
	Gr 10	16	Algebra I		Gr 11	48.4	35.8
Missouri		18.4	8.3				

Note: Prepared from data in state plans submitted to the U.S. DOE or obtained from state education departments between summer 2003 and spring 2004. Not all states have yet determined starting points for calculating AYP or we were unable to obtain information. Missing states had no data available. * VT and NY have calculated an index score rather than a percentage point. The DOE URL for state plans is <http://www.ed.gov/admins/lead/account/stateplans03/index.html>

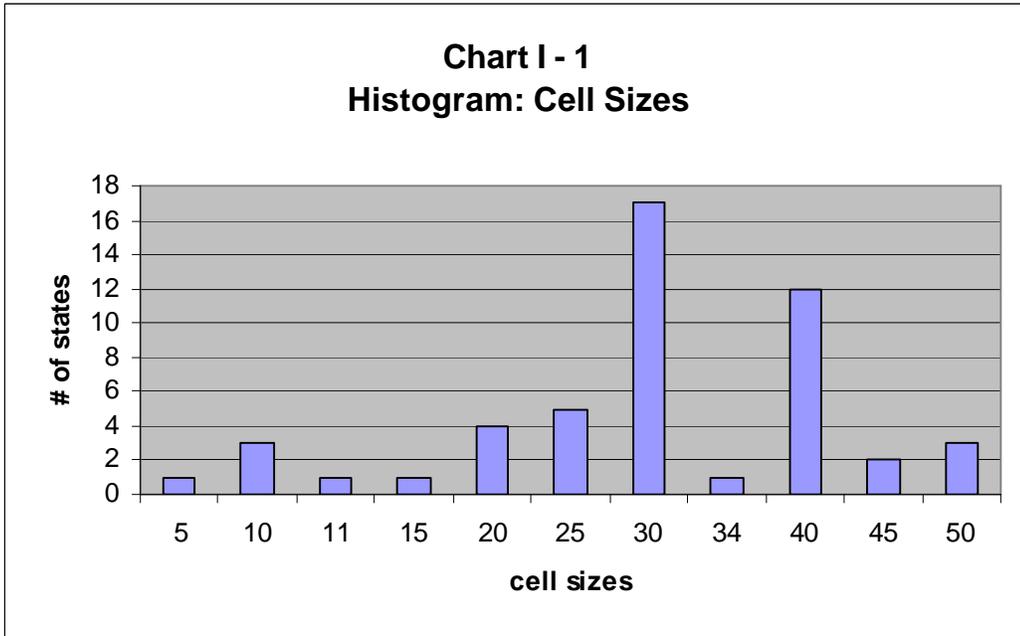
Table 1-3: State Cell Sizes: Minimum Number of Students to Be in a Cell

State	Min. # students to use subgroup in AYP	Min. # to calculate participation (if separate)	State	Min. # students to use subgroup in AYP	Min. # to calculate participation (if separate)
Montana	95% CI, one tailed		Vermont	30 over 2 years, plus 99% CI; small school review	
North Dakota	uses 99% CI		Nebraska	30, 45 for SPED	
Maryland	5 and test of statistical significance		Florida	30, except "small schools" having population <30 but >10	
South Dakota	10 and a 99% CI		Oklahoma	30 for all and regular ed;	40
Louisiana	10 with CI of 99% for AYP	40		52 for subgroups (99% CI); data aggregated across years to reach minimum of 30; might combine small schools scores	
Utah	10 and 99% CI	40	Wyoming	30	40
New Hampshire	11	40	Ohio	30 except 45 for SPED	40
Massachusetts (as of '04)	20 students/yr for 2 yr rating; 15 in any one year; SE=2.5 for >50; SE=4.5 for 20<X<50; won't report without 95% CI		Puerto Rico	30 except 45 for SPED	40
Minnesota	20	40	Idaho	34	
Alaska	20		Alabama	40	
New Jersey	0 and 95% CI		'02-'03		
Maine	20, will combine up to 3 yrs to reach min sample size; uses 95% CI	41	Connecticut	40	
Arkansas	25		Delaware	40	
DC	25		Georgia	40	
New Mexico	25		Illinois	40	
Nevada	25, combine years to bring schools to min		Mississippi	40	
Oregon	42 scores (21-28 students)		New York	40	
Kentucky	10 per grade with a min of 30		N. Carolina	40	
Arizona	30		Texas	40; 50 for sub-groups; small schools can average yrs	40
Colorado	30		Pennsylvania	40 will combine years to meet min.	
Indiana	30		S. Carolina	40, combine up to 3 yrs	
Kansas	30		Wisconsin	40; except 50 for SPED	
Michigan	30		Rhode Island	45	
Missouri	30		Tennessee	45	
Washington	30		Virginia	50	
Hawaii	30	40	W. Virginia	50	
Iowa	30	40	California	100 scores or 50 students	

Note to Table 1-3:

- CI = Confidence Interval
- SE = Standard Error
- SPED = Special Education/Disability

Chart 1-1: Cell Sizes



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. Urda (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>

III. False Promises & Flawed Policies: Navigating The NCLB Maze

“Secretary Paige says lack of capacity is no excuse for denying a transfer. Well, which kids will we put out on the street to make room? It’s too much what they’re asking us to do.”

— Toni Oklan-Arko, director of state and federal programs for West Contra Costa (CA) schools

“I don’t know of a school district contiguous to a low-performing district that has said yes. No money comes with the students. Choice then becomes a policy hoax.”

— Arnie Fege, Public Education Network

Two years into NCLB, it is possible to look at what is happening in schools and districts as they take their first steps toward interpreting and implementing the law’s voluminous fine print. The initial applications of the transfer and tutoring provisions already reveal educationally harmful developments, even as educators strive to make sense of and implement the law.

On the one hand, there is widespread support for the broadly stated goal of identifying individual students and groups who are not well served by their public schools and offering them better educational quality and outcomes. There is also strong support, particularly from historically underserved populations, for the concept of holding schools accountable for educating all students to high levels by focusing on the progress of these individual groups.

On the other hand, there is a rapidly developing consensus among those charged with implementing the law that they are being set up for failure. This is particularly true when it come to the specifics of the school choice provisions.

Parents are beginning to see the transfer and tutoring provisions less as practical solutions to real problems than as a kind of hoax being perpetrated on them and their schools. In high-poverty, minority-majority districts, there are few if any attractive alternatives even if parents want to pull their kids out and move them to an unfamiliar school. High-performing schools in neighboring communities are

There is a rapidly developing consensus among those charged with implementing the law that they are being set up for failure.

Parents are beginning to see the transfer and tutoring provisions less as practical solutions to real problems than as a kind of hoax being perpetrated on them and their schools.

tantalizingly out of reach because the law does not require them to accept out-of-district students. In most cities, magnet schools and schools with entrance score requirements are not available to those seeking to transfer. Thus, students often are only able to move from one low-performing school to another. Tutoring funds are padding the bottom lines of for-profit providers, diverting scarce resources from public to private hands and making it harder to provide services to all who could benefit. The law's requirements to radically alter the governance of schools that do not make "adequate yearly progress" (AYP), known as "restructuring," propose a set of steps that are, at best, unproven as tools to improve education.

Some parents who choose to have their children stay put see the law adding bureaucratic and logistical burdens to already struggling schools without actually providing something educationally valuable to needy students. It may well be a positive sign that so many parents seem to want to stick with their schools and see them improve rather than transfer their children to another school. This may simply reflect a desire to stay with a known quantity rather than risk the unknown. Research has shown that moving students from school to school itself damages academic progress. Or the reluctance on the part of parents to transfer their children may reflect an understanding that school ratings based purely on test scores do not provide an accurate picture. A parent's sense of a child's well-being can be a more meaningful and accurate gauge.

Lee Ann Early, a co-president of Peachtree Elementary School's Parent-Teacher Association in metropolitan Atlanta, wasn't surprised to learn that only one student chose to transfer under the law. "The parents who are at Peachtree know what a good school it is," Early said. "They know the phenomenal things that are going on there" (MacDonald, 2003a).

And Bruce Wilson, a black, middle class Chicago parent and PTA activist explained, "I think parents see through the illusion and the hype... [M]ost parents have evaluated their child's school based on factors beyond test scores, based on a wide range of needs... If the point is to 'leave no child behind,' then who's there for the children being left behind in failing schools?" (FairTest Examiner, 2002).

Some parents who choose to have their children stay put see the law adding bureaucratic and logistical burdens to already struggling schools without actually providing something educationally valuable to needy students.

"I think parents see through the illusion and the hype... [M]ost parents have evaluated their child's school based on factors beyond test scores, based on a wide range of needs... If the point is to 'leave no child behind,' then who's there for the children being left behind in failing schools?"

-Bruce Wilson

Yet the fear remains among both school officials and parents that when schools are labeled and sanctions applied, the failing tag could become a self-fulfilling prophecy and have dire consequences for all those who remain.

These developing patterns are troubling whether one focuses on the short or the long term. In the short term, implementation of the transfer and tutoring provisions has been characterized by chaos and confusion rather than a sense of progress toward a clear goal. Though NCLB defenders suggest that the complexities will ease once the early kinks are ironed out, it is far more likely that as the numbers of schools in need of improvement increase each year, it will become even more difficult for cash-strapped districts to comply with NCLB's mandates.

Meanwhile, the restructuring provisions loom before both schools and districts. As a series of unproven and demonstrably unhelpful provisions take hold, it is becoming clear why NCLB will leave many children behind.

A. The Transfer Shell Game

Under the transfer provisions of NCLB, a school receiving funding through Title I (the main part of NCLB) that has not made adequate yearly progress (AYP) for two years, must allow its students to transfer to another school in the district that *is* making AYP.

NCLB's transfer provision is designed to allow students to transfer out of schools that are labeled "in need of improvement" (INOI) into "better" schools, where they will presumably benefit. It is one of the so-called "parental choice" provisions of the law. In practice, it illustrates one way NCLB has failed to deliver on its alluring promise of an escape hatch for needy students "trapped" in failing schools.

The evidence to date highlights a series of key issues:

- Nationally, relatively few students are transferring out of schools labeled "in need of improvement."
- In districts where significant numbers have transferred, there are more reports of chaos, confusion, and dislocation for both sending and receiving schools than of successful outcomes.
- For students in many low-income, minority populated

When schools are labeled and sanctions applied, the failing tag could become a self-fulfilling prophecy and have dire consequences for all those who remain.

As a series of unproven and demonstrably unhelpful provisions take hold, it is becoming clear why NCLB will leave many children behind.

districts, there are few if any schools available to transfer into. Rural districts often have only one school. High-performing districts are not required to accept students from neighboring, low-performing districts.

- Local officials say they have insufficient time and resources to comply with NCLB provisions and are sometimes embroiled in disagreements with federal officials over which schools belong on the lists of INOI schools.
- Transfer costs are high and borne in many cases by already struggling schools and districts, making it more difficult for them to improve.
- As AYP demands increase and more schools land on INOI lists, it will become increasingly difficult for districts to comply with transfer provisions.
- The response from federal officials to questions from state officials about transfer requirements has been inconsistent, confusing and inadequate.

As result of all these factors, transfers are shaping up as a combination of a false promise to parents seeking better educational opportunities, a recipe for educational damage, and a lightning rod for parent and educator disapproval of the law.

Few Students Are Transferring

Across the country, it appears that few students are transferring out of schools dubbed “in need of improvement.” For a variety of reasons, including parents’ unwillingness to send children far away to unfamiliar neighborhoods where it would be hard to be involved in the new school, many parents have consciously chosen to have their children stay in their home schools, “failing” or not. Other parents say they weren’t informed of the option to move, or the information they received wasn’t clear. Some of those who *have* tried to transfer their children have found their options limited or nonexistent. Some districts maintained they have no space available. Others apparently created obstacles to transferring. Rural districts often have no second schools for students to transfer into, or the closest school may be many miles away. “Parents read about choice, then they find the doors barred,” said Arnold Fege, director of public engagement and advocacy for the Public Education Network (PEN) in Washington, DC (Gottlieb, 2003).

Transfers are shaping up as a combination of a false promise to parents seeking better educational opportunities, a recipe for educational damage, and a lightning rod for parent and educator disapproval of the law.

“Parents read about choice, then they find the doors barred.”

-Arnold Fege

Press reports provide snapshots like these: In New Orleans, of 35,000 eligible to transfer, 1,100 applied and 700 were turned down (Rasheed, 2003). In Warwick, Rhode Island, 375 parents were invited to attend a meeting on school choice and six showed up (Borg, 2003). In Gwinnett County, GA, only 60 students — less than half of 1 percent of the 17,000 who were eligible — were expected to transfer from the 14 county schools that were identified as “poor-performing” by the state Department of Education (Macdonald, 2003a). Of 51,716 letters sent to parents in six Bay Area districts where there were state-designated “schools in need of improvement,” just 1,018 students — fewer than 2 percent — requested transfers (Asimov, 2003).

NCLB proponents and other observers predicted an uptick in the number of transfers after early problems were worked out. Indeed, a survey conducted by the Council of the Great City Schools found three times as many students transferred in 2003-2004 than in the prior school year. Executive Director Michael D. Casserly attributed the increase to “time, experience and better communications with parents,” which he said may solve “some—though not all—of the problems districts are having with the choice provision.” Still, the total number cited by the survey for the 2003-2004 school year, 18,000, represents less than 2 percent of children eligible to transfer (Robelen, 2004).

The reasons for the non-responsiveness of parents are varied. In some cases, students transferred out and then transferred back when they felt unwelcome in their new school. Chicago is likely not the only city where students who transfer to a school under NCLB are reportedly referred to as “no child left behind kids,” suggesting they carry a stigma with them (PURE, 2003).

Some parents appear to share school officials’ worries that the transfer provision will set in motion a vicious cycle that will only undermine schools in which they have invested their resources, efforts and hopes for the future. For Mary-Powel Thomas, of Brooklyn, NY, talk about transferring is frustrating. She and other parents have worked closely with the teachers and principal at PS 38 to improve the school. She told a reporter she’s afraid that the label “failing” will cause many of the better students to leave, making it even more difficult in the future for the school to meet the new standards (Marks, 2003).

The total number cited by the survey for the 2003-2004 school year, 18,000, represents less than 2 percent of children eligible to transfer.

Some parents appear to share school officials’ worries that the transfer provision will set in motion a vicious cycle that will only undermine schools in which they have invested their resources, efforts and hopes for the future.

In addition, many schools to which parents can transfer their children are little better than the ones they are leaving. In Chicago, nearly 100 schools took in low-performing students from schools that had not made AYP for two years. Many receiving schools had average test scores only marginally higher than the schools students left behind. Thus, families involved with Parents United for Responsible Education have explained that they view the law as a fraud: They are not able to send their children to better schools, but the resources needed to improve their schools are also not available (PURE, 2003).

Public opinion polls have found that while parents may regard schools in general as subpar, they often give their own children's schools high marks. William G. Howell, a professor at the John F. Kennedy School of Government at Harvard University, surveyed public school parents in Massachusetts and found that 82 percent of parents gave their children's schools a grade of A or B. He said a reason why there are not more transfers may be simply that parents are satisfied with their schools and tend to think other schools are worse (Robelen, 2004).

Transfers Produce Chaos

In the few cities with significant numbers of transfers, such as New York, where 8,000 switched, students have found themselves in overcrowded schools lacking the space and staff to handle the influx. (New York City officials said only 8,000 out of the roughly 300,000 students in 315 schools who were eligible chose to transfer.) Principals said they received little notice that they would be receiving the students and no extra resources to provide for them.

New York Times education columnist Michael Winerip (2003) has been vividly illustrating the chaos ensuing in some New York schools as a result of transfers. One story highlighted students who wound up transferring from one school dubbed INOI to another "failing" school. The reason city officials deemed the second school eligible to receive the students, although it had failed to make AYP, was that it was not a Title I-funded school. (Schools that do not receive Title I funding must test and report results, but are not subject to NCLB sanctions.)

Winerip also interviewed Principal Louis Delgado of Vanguard School in Manhattan, which swelled to 440 students from 330. Winerip

In addition, many schools to which parents can transfer their children are little better than the ones they are leaving.

Thus, families involved with Parents United for Responsible Education have explained that they view the law as a fraud.

wrote: “The federal transfers scored the lowest level on state tests. Many are over-age, 16- and 17-year-old ninth graders. Mr. Delgado has no other administrators and instead used his resources to hire teachers and keep class size small. Until this year, he had 22 to a class; now it is 30. ‘We’ve had more fighting in one month than we did all last year,’ he said. ‘And there’s no extra resources. It destroys morale’” (Winerip, 2003).

NCLB provides little support for the planning that districts must do to implement transfers, but that planning takes substantial staff time and is expensive to the district. The non-partisan Center on Education Policy (2003), which is monitoring NCLB’s implementation, released a preliminary report in October 2003 titled *Implementing the No Child Left Behind Act*. The CEP said local officials in the 15 districts examined are “hopeful” they will be able to meet NCLB’s goals of 100 percent proficiency. Despite the report’s optimistic tone, its findings reflect the difficulty and frustration of implementing the transfer provision.

CEP summarizes its findings on school choice: “The NCLB requirements for school choice and supplemental educational services (tutoring) have been very time-consuming for some case study districts to implement and have been little used by parents, to date. In Grant Joint Union High School District, California, the director of categorical programs estimated that he and his staff spent more than 300 hours in 2002-03 implementing NCLB school choice and another 500 hours arranging for supplemental services in eligible schools.” Although large numbers of students were eligible to transfer in Grant, no students chose this option, despite the 300 hours invested by district administrators.

Among the harmful consequences of the NCLB transfer provision is its potential harm to the popular small schools movement, which has attracted great attention and substantial funding from The Gates Foundation and others. Even if there were adequate resources to absorb transfer students into some of these successful small schools in New York City and elsewhere, the transfer provision clearly strikes at the heart of the notion of building a small, stable, high-achieving community of students and teachers.

Disagreements with federal officials have also produced local confusion. In mid-August, Georgia administrators were still haggling

NCLB provides little support for the planning that districts must do to implement transfers, but that planning takes substantial staff time and is expensive to the district.

“In Grant Joint Union High School District, California, the director of categorical programs estimated that he and his staff spent more than 300 hours in 2002-03 implementing NCLB school choice and another 500 hours arranging for supplemental services in eligible schools.”

-Center on Education Policy

with the federal government over which schools would have to offer parents the transfer option. Federal education officials claimed Georgia misinterpreted NCLB and said that 270 more schools than the state had identified would need to offer transfers. State officials insisted that the 270 had made the state's testing goals during the 2002-03 school year and should not be on the list. But officials from the U.S. Department of Education said schools need to make their testing goals two years in a row before they can stop offering transfers or tutoring (Tofig, 2003). The battle left the districts, schools, teachers, parents and students uncertain of their situation and unable to plan for the coming year.

No Place to Go

Some early reports depicted a chaotic atmosphere in districts with so many schools dubbed INOI that there was no place for students to go if they chose to transfer, as was true in Illinois' Hazel Crest District 152 ½, where every school was deemed INOI. In downstate Illinois districts, such as Cairo Unit School District 1, officials would have had to bus students as far as 30 miles or into another state to find a school in compliance with NCLB's AYP dictates. "This law is ridiculous," said Robert Isom, superintendent of the Cairo district. "Does anybody in their right mind seriously expect me to move my kids into a Kentucky school?" (Banchemo and Cholo, 2003).

NCLB does not provide for interdistrict transfers except on a voluntary basis, and few school boards have welcomed students from other districts who are likely to lower average test scores. Arnold Fege of the Public Education Network said he hasn't heard of one high-performing district nationwide that has allowed children from low-performing districts to transfer in (Snyder, 2003). In Rhode Island, for example, the superintendents of Providence and Central Falls asked neighboring school systems to open up their schools but were told no (Borg, 2003). In California, the Grant Joint Union High School District sent letters to bordering districts asking if they would be willing to accept inter-district transfers, but these districts either did not have space for more students or chose not to accept out-of-district transfers (Center on Education Policy, 2003).

Chicago Mayor Richard Daley has become a prominent NCLB critic, calling the law a "logistical nightmare" and citing in particular the transfer provision, which he said is impossible to carry out with so

"This law is ridiculous," said Robert Isom, superintendent of the Cairo (IL) district. "Does anybody in their right mind seriously expect me to move my kids into a Kentucky school?"

In California, the Grant Joint Union High School District sent letters to bordering districts asking if they would be willing to accept inter-district transfers, but these districts either did not have space for more students or chose not to accept out-of-district transfers.

few available seats (Rossi, 2003b). In Chicago, 19,000 out of an estimated 250,000 eligible students requested transfers from failing schools, but school officials said there were only 1,100 seats available. *The Chicago Tribune* reported in September 2003 that only 481, or 44 percent, of those students who won those seats actually tried to claim them as of the first week of school. By April 2004, the odds of a student finding a transfer spot were worse than ever, with 190,000 elementary students eligible to vie for 500 seats. That means there is one seat available for every 380 eligible students, whereas last year it was one seat for every 247 students. There are no high school seats at all available for transferring students (Rossi, 2004).

The U.S. Department of Education's (DOE) regulations say the law "does not permit" schools "to preclude choice options on the basis of capacity constraints." Health and safety regulations and state laws that disallow school transfers are the only acceptable bases for denying student transfer requests.

Whether or not districts concur with the Department's claim that capacity does not matter, some states have class-size limits. Florida voters, for example, approved a Constitutional amendment in 2002 setting ceilings on the number of students in each classroom (Richard, 2004). NCLB presumably does not override such legal requirements. DOE tried to argue that NCLB transfer requirements trump union contracts with class size provisions. However, the federal law expressly states that it does not supercede either current or future contracts, so the DOE was forced to back down.

DOE also claims that NCLB supercedes federal court desegregation orders, leading some critics to wonder if this was not part of a Bush Administration effort to end school desegregation. One civil rights attorney observed that NCLB "cannot trump the Fourteenth Amendment."

Federal officials have called on Chicago to try harder, suggesting for example that they purchase trailers to create more seats at higher performing schools, offer classes via the Internet, or create incentives for higher performing districts to take underperforming students. Chicago Schools CEO Arne Duncan called the trailer idea "interesting," but noted that the cost, \$500,000 to \$1.2 million per trailer, would again be borne by the district (Rossi, 2003b).

By April 2004, the odds of a Chicago student finding a transfer spot were worse than ever, with 190,000 elementary students eligible to vie for 500 seats.

DOE also claims that NCLB supercedes federal court desegregation orders, leading some critics to wonder if this was not part of a Bush Administration effort to end school desegregation.

Mayor Daley is far from alone in his dim view of the transfer provision, which has sparked anger even from supporters of the law, such as Maryland education Superintendent Nancy Grasmick, who referred to the requirements as “a nightmare.” Los Angeles Superintendent Roy Romer said he saw no way to comply as Los Angeles is already 200,000 seats short and using its classrooms on staggered, year-round schedules (FairTest *Examiner*, 2003).

Suggestions such as those given by federal officials to Chicago call into question whether anyone is seriously thinking through the consequences of such remedies. Would only the “No Child Left Behind” students be housed in trailers, potentially adding to the stigma they would bring with them from their “failing” school and neighborhood? Or would the trailers house a mix of old and new students, and how would parents of kids already in the higher performing school react to their kids being taught in trailers as a result of the influx? There is no evidence to support the notion that suddenly putting a group of children in trailers on the grounds of a successful school will lead to academic improvement.

Chicago’s situation reveals the ways in which the transfer option fails to provide real alternatives to parents and students and can even exacerbate inequality. In making schools available for transfers, “elite” schools – magnet schools or those requiring test score cut-offs or other special admissions requirements – are typically exempt under NCLB provisions. The schools that remain are usually marginally better, if at all.

Winerip’s columns in *The New York Times* and other evidence suggest similar problems in New York City. There, gifted and talented programs, which are overwhelmingly white, Asian and upper middle class, are exempt. One family in New York with two children in a public exam school for gifted and talented children noticed that their school was exempt from accepting NCLB transfer students, yet another public school in the same building had as many as 48 students per class (Kaplan, 2003).

The massive increases in class sizes run counter to research findings that small class size does matter. The well-known Tennessee STAR project found that small class sizes did boost test scores and particularly benefited low-income children. Similarly, research on Wisconsin’s

There is no evidence to support the notion that suddenly putting a group of children in trailers on the grounds of a successful school will lead to academic improvement.

One family in New York with two children in a public exam school for gifted and talented children noticed that their school was exempt from accepting NCLB transfer students, yet another public school in the same building had as many as 48 students per class.

SAGE program, designed to give a boost to young children from low-income families, has found small class sizes led to significant learning gains and helped close racial achievement gaps (Smith, Molnar & Zahorik, 2003). Thus, the new federal mandates that are intended to induce improved learning are sabotaging educational practices that have scientific evidence to show they work.

Finally, most rural schools will find no feasible options available for fulfilling the NCLB transfer provision. In many cases, rural districts have only one school. Other large districts may have schools many miles apart, leading to very lengthy bus or car rides and great expense.

Scarce Funds Diverted from Needy Schools

NCLB requires that at least 5 percent and up to 20 percent of a district's funds are to be used for transportation of transferring students and to cover tutoring expenses. If Title I support money is not transferred along with the students, the receiving schools face larger class sizes and a need for new books and supplies without any increase in funding. If funds do follow transferring students, the school that had been labeled INOI finds itself with reduced funding at a time when it is required to improve.

In Duval County, Florida, when 440 students transferred at the start of the 2003-04 school year, each student took \$5,268.87 in state and federal funding to the new school, meaning more than \$2.3 million in funding was redirected. A single school in the county, Eugene J. Butler Middle School, which lost 196 students, took a \$1 million hit in Title I funding. The result was the loss of eight teachers, an assistant principal and several teaching assistants (Garza, 2003).

Whether or not Title I funds move with students when they transfer, there is a serious question about who is served by such choice provisions. Since transportation costs are taken out of the district's Title I funding, money that could have gone into hiring teachers, providing professional development or other improvements is spent on busing. Whether money moves with students or not, sending and receiving schools face unpredictable funding levels and must spend scarce administrative resources on implementing transfers rather than improving schools. The end results are erratic and often of dubious benefit to even the few who choose to move.

Finally, most rural schools will find no feasible options available for fulfilling the NCLB transfer provision.

Since transportation costs are taken out of the district's Title I funding, money that could have gone into hiring teachers, providing professional development or other improvements is spent on busing.

Some analysts question whether there is any evidence that transfers will improve educational outcomes. Research does indicate that low-income children in middle-income schools do better than low-income children in low-income schools, but large-scale transfers of the sort encouraged by NCLB are a new, unstudied phenomenon.

Some observers predicted that if successful schools were sent the lowest performers from neighboring schools, the transferring students could end up lowering the receiving school's test scores. Reduced test scores could then prevent the previously successful school from achieving AYP, making it a school in need of improvement.

This worry became reality in Chicago in 2003, according to a *Chicago Sun Times* report. The article found that Chicago schools that took in NCLB transfer students were more likely to see drops in reading and math scores than the system as a whole. The effect was most severe in schools taking between 15 and 29 students. Ninety-two percent of these schools saw reading scores fall, compared with 69 percent of all schools systemwide. In math, 75 percent of schools receiving 15 or more students experienced a drop, while 39 percent of schools decreased systemwide (Grossman, 2003).

Fewer and Fewer Choices

Some observers have predicted the chaos will abate as districts get themselves better organized and parents learn more about their options. But more and more schools are likely to be labeled "in need of improvement" (INOI) each year, thus leaving fewer schools eligible to accept transfers. In most large cities, the great majority of schools that receive Title I funds will soon be INOI and not eligible to receive transfers.

Perhaps because intradistrict transfer options are few and will continue to diminish as lists of schools INOI grow, U.S. Education Secretary Rod Paige suggested in an online chat sponsored by *Education Week* (2003) that there may be more leeway than the law and regulations appear to allow. Asked what districts should do if all non-failing schools in a district are at capacity, Paige replied: "The school district can come up with innovative ways to meet the students' needs. They can use supplemental services sooner, can create charter schools, and can create schools inside of schools. They can even use technology to provide choices and options for students. The district can also work

Research does indicate that low-income children in middle-income schools do better than low-income children in low-income schools, but large-scale transfers of the sort encouraged by NCLB are a new, unstudied phenomenon.

More and more schools are likely to be labeled "in need of improvement" (INOI) each year, thus leaving fewer schools eligible to accept transfers.

with other nearby districts to provide more options to parents. All of these choices can be accomplished by effectively utilizing Title I funds. The bottom line is these students need extra assistance to attain the education they deserve.”

Far from offering clear guidance to confused and frustrated school administrators, Paige’s response raised a host of new questions about just how much leeway the law would allow, for example, on the issue of whether tutoring could be offered instead of transfers. His comments did appear to open the door to gaming the system. For example, if a school is divided in two and each half given a new name (an increasingly common practice in Boston, for example), it may well be that neither would start out INOI.

A Guaranteed Failure

It is becoming increasingly clear that the transfer policy is not about improving education. “What this does is send a few kids out and leave all the rest behind,” said Madeline Talbot of Chicago ACORN, which has opposed use of transfers and argued for using the money in the lowest performing schools (FairTest *Examiner*, 2003).

“It’s not the only way to go about it,” explained Melinda Anderson, spokeswoman for the National Education Association. “There needs to be a more comprehensive look at this as opposed to the silver bullet... Giving parents the option to transfer to a ‘higher achieving school’ looks good on paper, but the reality of the situation is that it undermines the responsibility of school districts to make every school a high achieving school” (Anderson, 2004).

Richard Elmore, professor of education at the Harvard Graduate School of Education, characterizes NCLB as “policymaking by remote control.” Writing in *Education Leadership* (2003), Elmore charged that some NCLB provisions may sound sensible inside the Beltway but not to those trying implement the law. “Parental transfer rights, if they work, increase instability in enrollments in low-performing schools and adversely affect the distribution of students among schools, without necessarily improving instructional practice in either the sending or receiving schools... Policy gets made in one place and implemented in another; how it gets implemented is someone else’s problem.”

“What this does is send a few kids out and leave all the rest behind,” said Madeline Talbot of Chicago ACORN.

“Giving parents the option to transfer to a ‘higher achieving school’ looks good on paper, but the reality of the situation is that it undermines the responsibility of school districts to make every school a high achieving school.”

-Melinda
Anderson

The primary argument for transfers is that they will enable students to “escape” low-performing schools. The policy may work for an extremely limited number of students. But as a means to improve educational opportunities for all children, transfers will not succeed for the following reasons:

- In districts with many low-income children, there are or soon will be too few schools to which students can transfer.
- Schools that are not required to accept transfers have no incentive and many disincentives to do so.
- Many higher-scoring schools and programs are “off limits,” often to preserve enclaves of relative privilege.
- Suggested options, such as allowing overcrowding or using trailers, will only diminish the quality of education at schools that accept transfers.
- Students who transfer may go from a school that has programs for limited English students or for students with disabilities or other specific needs to a school that has no programs to meet those needs.
- The chaos of continually moving students, of trying to adjust the teaching force to match classroom populations, and of trying to re-design the process will contribute to continued educational problems and to burnout.
- Parents increasingly see transfers as an illusion, even a trick, and refuse to participate.

B. Tutoring: No Test Prep Company Left Behind

NCLB’s supplemental services provision represents what remains of a failed push for a federal voucher program by President Bush and his allies. Under the law, low-income students in schools that do not make AYP for three years must be offered supplemental services, or tutoring. The tutoring can be in reading, language arts, and math, before or after school or on weekends, by providers approved by state departments of education. Although public school districts themselves can be approved to provide tutoring, a wide array of test coaching companies has jumped into this field and seems to be the constituency most enthusiastic about this provision.

High-quality tutoring is widely understood as educationally beneficial. Whether its implementation under NCLB will produce powerful benefits remains to be seen, but for many reasons such success

Schools that are not required to accept transfers have no incentive and many disincentives to do so.

Although public school districts themselves can be approved to provide tutoring, a wide array of test coaching companies has jumped into this field and seems to be the constituency most enthusiastic about this provision.

is unlikely. Tracking the implementation of the tutoring provision reveals:

- The supplemental services provision is based on a set of faulty assumptions about the reasons for low student achievement, which result in a faulty set of solutions.
- Private firms that provide tutoring and/or test preparation services stand to reap large gains from the law, while already strapped districts see their budgets pinched further and have to cut back on basic services to all students.
- The diversion of public funds to costly private tutoring firms is not cost-effective and means that fewer students overall can have access to tutoring services.
- As with the transfer provision, few parents have taken advantage of tutoring options.
- Tutoring focused primarily on raising test scores, as called for by the law, does not constitute the kind of quality instruction that struggling students need and deserve.
- Few supplemental service providers are prepared and willing to accommodate special needs and limited English proficient students.
- The law's assumptions lead to absurdities such as computer-based tutoring being judged to be of a higher quality than one-on-one tutoring by certified public school teachers.
- Large urban districts in particular are struggling to comply with contradictory mandates with inadequate resources.

The tutoring provision is another example of NCLB offering inadequate “choices” for a small minority of students rather than providing sufficient resources and support to improve public schools for all students.

Faulty Assumptions Make for Faulty Solutions

The logic of NCLB's tutoring provision rests on a particular set of assumptions, in addition to the simple one that good tutoring often helps, including:

- Low test scores are primarily due to poor instruction by public school teachers.
- The threat of having to devote resources to provide or pay for supplemental services will spur schools to improve their practices, boost test results and avoid the penalty.
- Outside tutoring services can compensate for poor instruction

The supplemental services provision is based on a set of faulty assumptions about the reasons for low student achievement, which result in a faulty set of solutions.

The tutoring provision is another example of NCLB offering inadequate “choices” for a small minority of students rather than providing sufficient resources and support to improve public schools for all students.

- by improving test scores, an educationally valuable service.
- Offering parents this kind of “choice” is inherently valuable.

Meeting the challenges faced by districts trying to implement NCLB’s supplemental services mandates might be worthwhile if there were evidence that such assumptions were valid. However, there is substantial evidence that calls these assumptions into question.

Consider the assumption that teachers of low-income students (whose schools are most likely to not make AYP) don’t do as good a job as teachers of more affluent children. Researchers have found that while low-income, middle-class and upper-income students all make comparable academic gains during the school year, low-income students fall behind from year to year primarily because they tend to lose ground during summer vacations. Higher income students, on the other hand, tend to gain ground during the summer. While some have used these findings to push for year-round schooling, the authors of one study point out that it is not traditional schooling that accounts for the gains made by higher income students during the summer.

On the contrary, according to researchers Karl Alexander, Doris Entwisle and Linda Olson of the Johns Hopkins University, higher income students make gains because of the kinds of nonacademic enriching activities their parents can afford to provide during summer vacations (Alexander *et al.*, 1997). “We found that better off children in the [study] more often went to city and state parks, fairs, or carnivals and took day or overnight trips. They also took swimming, dance, and music lessons; visited local parks, museums, science centers and zoos; and more often went to the library in summer” (Bracey, 2003).

Education researcher Gerald Bracey (2003) says this evidence suggests that rather than causing student failure, public school teachers who work with low-income students are preventing failure during the school year. But even the best teachers cannot control the fact that their students’ lack of access to enriching activities causes them to fall farther and farther behind their more advantaged counterparts each summer.

If summer loss, rather than inadequate school-year instruction, explains most of the growing achievement gaps (at least as measured by test scores), a policy that favors outside providers and test prepara-

Low-income students fall behind from year to year primarily because they tend to lose ground during summer vacations..

Education researcher Gerald Bracey (2003) says this evidence suggests that rather than causing student failure, public school teachers who work with low-income students are preventing failure during the school year.

tion companies seems wasteful and ill-advised. Of course, this does not mean that tutoring itself is a bad idea, particularly during the summer months. Rather, it suggests that shaping a tutoring program based on the premise that the school's teachers are not effective is not wise.

Under NCLB, a school not making AYP cannot spend money on in-house tutoring, presumably because its teachers are not doing a good enough job. Completely uncertified instructors in a private company, however, may be paid for tutoring at another location. That location may be relatively inaccessible to many students and families. Thus, the flawed assumption about teacher quality leads to often more expensive privatization and lack of access for students who do not have transportation to private tutoring locations.

Gold Linings for the Sylvans

Large companies such as Sylvan Learning Centers and Kaplan Inc. could reap a windfall from the provision, which is expected to pad the profits of an industry that is worth approximately \$8 billion and growing rapidly (Clowes, 2003). The U.S. Department of Education (DOE) estimates that NCLB will pump an additional \$1 billion per year of public funds into tutoring.

Wall Street analysts see many lucrative opportunities in NCLB for textbook publishing, test development and scoring, and teacher training, but none greater than that for tutoring companies. "We've heard the supplemental education provision jokingly referred to as the 'Sylvan clause,'" said Jeffrey M. Silber, a senior vice president with Gerard Klauer Mattison, a New York City-based investment bank that follows education stocks. "Sylvan is the number-one player in the tutoring market, and the law is just a huge windfall for it" (Walsh, 2003).

Sylvan boasts that it has been approved as a NCLB supplemental educational services provider in 26 states and the District of Columbia, more than any other company. In celebrating its good fortune in September, Sylvan hosted President Bush and Education Secretary Rod Paige at a center in Nashville. As public school districts nationwide lay off teachers to cope with tapped-out school budgets, Sylvan is recruiting regional managers to coordinate all the business it expects to come its way via NCLB (various Sylvan online employment ads, 2003).

Large companies such as Sylvan Learning Centers and Kaplan Inc. could reap a windfall from the provision, which is expected to pad the profits of an industry that is worth approximately \$8 billion and growing rapidly.

As public school districts nationwide lay off teachers to cope with tapped-out school budgets, Sylvan is recruiting regional managers to coordinate all the business it expects to come its way via NCLB

Not to be outdone, the Princeton Review test preparation company reported in November 2003 that its K-12 Services division had a 74 percent increase in revenue from the prior year. According to a press release, “Key highlights included the successful launch of the Company’s interim assessment testing program for 500,000 students in New York City; the extension of its partnership with Philadelphia; and, two pilot programs in Virginia that include a benchmarking program for approximately 1,400 students and an online tutorial to help high school students pass the required state reading exit exam” (Princeton Review, 2003).

More Money for Fewer Services

Philadelphia schools chief Paul Vallas has complained that the tutoring provision forces him to pad the profits of outside providers while hampering his ability to serve the largest number of students in need. Vallas claimed he would be able to serve 27,500 more students by mandating that they obtain tutoring through the district’s extended day program than would be served by outside providers. The district’s extended day tutoring would cost \$300 per student for 120 hours of instruction, but outside providers could receive up to \$1,800 per student for as little as 30 hours of instruction. “I’m not averse to outside providers, but to spending \$1,800 per student,” Vallas said. “My frustration is that there are so many kids in need, and my frustration is accountability. I’m just trying to serve the most children possible with the highest-quality program” (Mezzacappa, 2003).

Supplemental services providers in Philadelphia, including for-profit firms, nonprofits, individuals and faith-based groups, could earn more than \$15 million a year for tutoring (Mezzacappa, 2003).

Similarly, one news report described the painful irony of a Chicago district that had to cut down on the number of children receiving after-school tutoring to pay for a smaller number of students to be tutored by Sylvan (Lutton, 2003).

Will Parents Buy It?

Some parent groups have either not been impressed with the tutoring provision’s implementation or are simply not interested. Complaints include claims that districts are either not yet providing supple-

Philadelphia schools chief Paul Vallas has complained that the tutoring provision forces him to pad the profits of outside providers while hampering his ability to serve the largest number of students in need.

Similarly, one news report described the painful irony of a Chicago district that had to cut down on the number of children receiving after-school tutoring to pay for a smaller number of students to be tutored by Sylvan.

mental services or not communicating effectively to parents about their availability.

Lack of transportation to outside providers could also be hampering students' access to tutoring services. In New York, for example, the city's Department of Education said only 30,333 children requested tutoring in the 2002-2003 school year, 12.5 percent of the 243,249 eligible. Contrary to predictions that more parents would take advantage of tutoring opportunities once first-year kinks were ironed out, even fewer parents initially availed themselves of tutoring options in 2003-2004. An October 2003 *New York Times* article reported, "Despite stepped-up efforts by the city to publicize the tutoring program this year, some of the largest private providers of tutoring services said yesterday that enrollment numbers were trailing last year's." The city school department responded by extending the deadline for parents to apply (Herszenhorn, 2003). Perhaps as a result of the extension and more vigorous outreach efforts, more eligible parents learned of the tutoring option, so that by November, 40,369 youngsters, or 19 percent of those eligible, had signed up for tutoring (Gross, 2003).

Parents in New York and other cities complained that information on tutoring is hard to come by and confusing when available. In January 2003, two parents filed suit against the New York and Albany public schools, saying they had been denied the right to transfer out of failing schools or obtain tutoring. Albany Superintendent Lonnie Palmer, however, said the lawsuit was politically motivated, with obvious support from groups pushing for charter schools and vouchers (New York State United Teachers, 2003). The case was dismissed in June 2003 by a federal district judge who said the law does not confer any such rights that can be enforced in court (Walsh, 2003).

Choice Provisions Divert Resources from School Improvement

The Association of Community Organizations for Reform Now (ACORN) has concluded that the money made available for tutoring—5 to 15 percent of a district's Title I budget—is not adequate to meet the need. In November 2002, ACORN reported on states' compliance with NCLB's supplemental services and other provisions and found most states lacking. The report said that three-fourths of the 23 states it studied were not yet providing supplemental services. Los Angeles, for example, had 150,000 eligible students but had funding for only

Lack of transportation to outside providers could also be hampering students' access to tutoring services.

The Association of Community Organizations for Reform Now (ACORN) has concluded that the money made available for tutoring—5 to 15 percent of a district's Title I budget—is not adequate to meet the need.

44,000 to actually receive services. Baltimore had funds to serve only one-third of its 18,000 eligible students.

A subsequent ACORN report, released in May 2003, noted, “The Bush administration has pushed states to implement NCLB provisions related to school choice, supplemental services, and academic testing. The choice and supplemental service provisions uproot students and take money out of school district funds, which could be used to train and retain more qualified teachers” (ACORN, 2003).

What Kind of Tutoring?

A critical question is what is the nature and quality of tutoring services being provided to those relatively few students who are getting them. The law specifies that supplemental services must be designed to increase the academic achievement as measured by the state’s assessment system, must enable these children to attain proficiency in meeting state standards, and must be “of high quality and research-based.”

However, the American Evaluation Association (AEA) recently challenged NCLB’s criteria for “research-based” methods, charging that the emphasis on experimental design will deny other “scientifically credible evaluation methods” and result in fruitless expenditures on some contracts while others go unevaluated (American Evaluation Association, November 2003).

In the case of supplemental services, the emphasis appears to be on getting providers with a track record of raising test scores, or in other words, those whose expertise is in test preparation. An *Education Week* article on the interest of test preparation firms in the tutoring provision pointed out that there has traditionally been a distinction between tutoring such as that provided by firms like Sylvan and test preparation of the sort offered by Kaplan and Princeton Review, but that the language of the law clearly indicates that firms offering purely test preparation would qualify. *Education Week* quoted John Katzman, the founder and chief executive officer of Princeton Review. “For the kids who are failing in those schools, we’re not the guys,” he said. “We’re not a remediation company.” But, he added, “There are plenty of kids who fail state tests by 3 percent, and it’s because they just don’t know how to take tests.” Therefore, he said, his company

In the case of supplemental services, the emphasis appears to be on getting providers with a track record of raising test scores, or in other words, those whose expertise is in test preparation.

“For the kids who are failing in those schools, we’re not the guys,” he said. “We’re not a remediation company.”

-John Katzman

might work with another company offering services under the tutoring provision (Walsh, 2003).

Teacher unions are angry over the government's insistence that tutors do not have to be certified educators, arguing that if the federal government insists that all teachers be "highly qualified," which effectively means certified, tutors should be as well. The Department of Education says parental choice is more important.

Few Providers for Students Most in Need

Several recent studies have documented the scarcity of supplemental service providers that are prepared for and willing to accommodate special needs and limited English proficient students. For example, a survey in New York done by Advocates for Children (2003) found most providers had no services designed for those groups. Among the study's findings were that nearly 60 percent of responding providers stated they did not have specific services for English language learners (ELL), and 44 percent did not know whether or not they were actually providing services to ELL students.

Advocates for Children also reported that though disabled students make up 11 percent of the New York school population, just 3.5 percent obtained supplemental services through NCLB last year. According to Advocates for Children executive director Jill Chaifetz, not only were few served, but it appears that most private providers were unprepared to serve them effectively. The study found that 44 percent of providers said they did not have services available that accommodated the specific needs of disabled students, and 50 percent did not know whether any of the students they were serving were disabled.

A study by the Council of the Great City Schools said providers often ask to be exempt from serving these more challenging groups of students. *Education Week*, reporting on the tutoring provision, quoted Ethel J. Collier, Chicago's director of supplemental services, who said these most challenging students had few options among providers. "In my experience, private providers have a very limited capacity to provide services for [English as a second language] and special ed students," said Collier. "Some kids they accepted, they later had to tell us they couldn't serve" (Gewertz, 2004).

Several recent studies have documented the scarcity of supplemental service providers that are prepared for and willing to accommodate special needs and limited English proficient students.

The study found that 44 percent of providers said they did not have services available that accommodated the specific needs of disabled students, and 50 percent did not know whether any of the students they were serving were disabled.

Computers More High-Quality than Teachers?

According to the logic of NCLB's supplemental services provision, teachers in schools that have failed to meet AYP mandates are not qualified to provide tutoring to struggling students, but computer-based programs fit the bill. For example, a school that failed to make AYP because one special needs student too few was tested will be ineligible to offer tutoring, but parents of low-income students in the school will receive a list of approved providers that includes an array of computer-based companies, for example, EdSolution, EduCare and SMARTTHINKING (Miner, 2003).

Barbara Miner (2003), in an article in *Rethinking Schools*, points out one obvious flaw in this logic. "There's one problem with some of the cyber-based companies, however. You need an up-to-date computer. Lightspan, headquartered in San Diego, has been particularly creative in resolving this dilemma. Its boasts that its "interactive standards-based program ... runs on the PlayStation PS-1 console, so it can go home with every student, where it helps increase family involvement."

There is an exception, however, to the regulation disqualifying schools that have been labeled "in need of improvement" from providing tutoring. Disabled students may be tutored by a school labeled INOI if there are no other providers with the capacity to offer them the accommodations spelled out in their individualized education plans. The regulations state that if the "identification for improvement status was the performance of its disabled or limited English proficient on assessments, then it would be *preferable* for the [local education agency] to serve those students through a contractor rather than by serving them directly." Preferable, but not mandatory, most likely because private contractors set up to meet disabled students' learning needs are difficult to find.

Large Urban Districts Struggle to Comply

Many big city school officials say they are struggling to comply with NCLB's sometimes contradictory mandates without sufficient funds to carry them out. The *Chicago Tribune* reported (Rado and Olszewski, 2003) that Chicago Public Schools (CPS) officials said they could afford to provide reading and math tutoring for only 25,000 or

According to the logic of NCLB's supplemental services provision, teachers in schools that have failed to meet AYP mandates are not qualified to provide tutoring to struggling students, but computer-based programs fit the bill.

Many big city school officials say they are struggling to comply with NCLB's sometimes contradictory mandates without sufficient funds to carry them out.

30,000 out of 133,000 eligible students in the 2003-2004 school year. CPS officials said there was insufficient federal funding to provide tutoring to the remaining 100,000, or 80 percent of those eligible. The *Tribune* said 231 city schools were required to offer special tutoring, and 133,000 notification letters were sent to parents. With a budget of 20 percent of Chicago's Title I funds, students were to receive at least two hours a week of tutoring from October through May at about \$1,500 per student. Depending on the program, tutoring can cost between \$20 and \$65 an hour.

CPS had argued that they should be able to first use their limited Title I funds to provide supplemental services in school, rather than spend money complying with the transfer provision. CPS Chief Executive Officer Arne Duncan pointed out that many of the schools mandated to allow students to transfer because they had not made sufficient AYP were actually making substantial progress, so the money would be better spent on tutoring than on busing kids from one school to another. "It just makes sense educationally," he said. "Where we're seeing improvement, we want to invest in the schools - not in more yellow school buses."

When eligible families were informed of the tutoring option, CPS said almost 90 percent of eligible parents failed to claim services for their children. According to the *Chicago Sun Times*, "only 14,931 of 133,000 low-income parents — or 11 percent — responded to Aug. 29 letters offering their kids extra help from 10 private tutoring companies, including such brand names as Sylvan and Kumon, or from their own schools" (Rossi, 2003a).

Duncan claimed that there had been a vigorous effort to inform parents of the option, but since so few responded, the schools should be able to decide how to spend \$20 million in unclaimed federal funds. Others said parents didn't respond because they didn't receive the information, didn't understand it, or lacked transportation to tutoring services (Rossi, 2003a).

When New Orleans offered tutoring to 7,500 eligible students in 16 schools identified as INOI, only 489 signed up. Parents had been notified by letters that seven state-approved providers—both for-profits such as Sylvan and Kumon Math and Reading Centers and nonprofits such as a YMCA—would offer after-school sessions at least twice a week, with some offering weekend sessions.

CPS had argued that they should be able to first use their limited Title I funds to provide supplemental services in school, rather than spend money complying with the transfer provision.

"Where we're seeing improvement, we want to invest in the schools - not in more yellow school buses."

- Arne Duncan

According to the *Times-Picayune*, city officials had been threatened with the loss of \$30 million in federal funds because the city had failed to offer tutoring to students in 12 schools that had been eligible since January 2003. The city responded by offering tutoring immediately to all students who requested it (Rasheed, 2003).

A fraction of those eligible for tutoring in Gwinnett County, GA, near Atlanta — 250 out of 3,000 — signed up to take advantage of the option. The district received \$477,000 in federal funds for tutoring and will pay a variety of providers \$1,230 per student for their services (Macdonald, 2003a).

Conclusion

If there was strong evidence that after-school and weekend tutoring would benefit struggling students, why not allocate resources to allow public schools to supplement their programs with these services, rather than impose it as a punishment for schools deemed “in need of improvement”? If education resources were not going to pad private company profits and duplication of space, schools might be able to provide better services to more students at lower cost.

C. Restructuring: No Quick Fixes

NCLB requires a school or district that has failed to make Adequate Yearly Progress (AYP) for five years to set up an alternative governance structure. This is the final stage of a three-step process that includes improvement, corrective action, then restructuring. Schools that lag behind in AYP for two years in a row enter an improvement stage that requires them to write and implement school improvement plans as well as implement the transfer option. After three years of not making AYP, schools enter the corrective action stage, which entails intensified district intervention as well as the tutoring option. After five years of inadequate progress, they move into the restructuring phase, also under district supervision. School districts that fail to make AYP are subject to a similar staged approach, also ending in restructuring. States are responsible for overseeing districts as a whole, identifying those needing improvement, and taking corrective actions when necessary.

The law specifies a set of “alternative governance arrangements”

A fraction of those eligible for tutoring in Gwinnett County, GA, near Atlanta — 250 out of 3,000 — signed up to take advantage of the option.

If education resources were not going to pad private company profits and duplication of space, schools might be able to provide better services to more students at lower cost.

for persistently failing schools. Districts may choose to:

- Reopen the school as a public charter school.
- Replace all or most of the school staff, including the principal.
- Enter into a contract with an entity, such as a private management company, with a demonstrated record of effectiveness to operate the school.
- Have the state directly manage the school.

In addition to these specific alternatives, there is one open-ended option: “Any other major restructuring of the school’s governance arrangement.” This option may have been added as an implicit acknowledgment that there is little evidence that any of the other alternatives are consistently successful.

The National Education Association estimates that already some 6,000 schools have not made AYP for two or more years and are in the stage of having to write improvement plans (see Table I-1) (NEA, 2004). As these schools are already behind, will receive limited new resources, and will have to both catch up to other schools and make additional progress, it is all too likely that most will soon be facing reconstitution, with many thousands more following soon after. Already, states which began to keep track of progress before NCLB was implemented have some schools that will face restructuring in the 2004-05 school year.

Charter Schools

While the number of charter schools has grown rapidly in recent years, they have not been used as a means of reconstituting existing “failed” schools. The record of student achievement in charter schools is decidedly mixed and the subject of ongoing debate. Certainly, there is no consensus that charters offer an inherently superior alternative to traditional public schools. There are studies that point to some schools making modest gains in standardized test scores (Miron *et al.*, 2002) or charter schools besting their regular public school counterparts in reading but not math, for example (Solomon *et al.*, 2001). One study of California charters found low-income students performed better in charter schools than in regular public schools (Slovacek *et al.*, 2001).

On the other hand, a study of New York charters called for a moratorium on charters in New York after finding no evidence that the schools improved student achievement (New York State United Teach-

The National Education Association estimates that already some 6,000 schools have not made AYP for two or more years and are in the stage of having to write improvement plans

Certainly, there is no consensus that charters offer an inherently superior alternative to traditional public schools.

ers, 2003). Similarly, a Western Michigan University study found traditional public schools did better than a group of Michigan charter schools (Miron and Horn, 2000). In 2002, the authors of a study of North Carolina charters also called for a moratorium after students in the schools fared worse in reading, writing and math than regular public school students (North Carolina Center for Public Policy Research, 2002).

Replace School Staff

Replacing staff is probably the most commonly used approach to reconstitution, but it has generally not succeeded as a means of improving schools.

A report titled “Can Failing Schools be Fixed?” produced by the conservative, pro-NCLB Fordham Foundation, looked at 20 kinds of interventions into failing schools, including what it called “reconstitution,” defined as replacing most or all of a school’s staff (Brady, 2003). The report’s conclusion was that some turnaround efforts have improved some schools, but success is not the norm; no particular intervention appears more successful than any other; interventions are uneven in their implementation and always hard to sustain; and it is nearly impossible to determine which interventions are most effective because they are attempted in very different situations.

The Fordham study looked at the 1997 reconstitution of six Prince George’s County, Maryland schools. Author Ronald C. Brady concluded the results “mirrored those of efforts in other locations where it has been implemented—effective in improving student achievement in some schools, but yielding little or no improvement in others” (Brady, 2003).

Similarly, Kent Peterson of the University of Wisconsin found mixed results, from which he drew several lessons: reconstitution is enormously complex and difficult; achievement results vary; it takes enormous resources and skill; and districts must consider unintended consequences like low teacher morale and political conflict (Peterson, 2000).

In a 2001 research paper on reconstitution, Mary Levy, of the Washington, DC, Lawyers’ Committee for Civil Rights Under Law,

In 2002, the authors of a study of North Carolina charters also called for a moratorium after students in the schools fared worse in reading, writing and math than regular public school students.

Replacing staff is probably the most commonly used approach to reconstitution, but it has generally not succeeded as a means of improving schools.

comments on the underlying assumptions of reconstitution, which mirror those that undergird much of NCLB. Proponents believe reconstitution will work because they believe “that school staff are the principle source and principle solution for poor school performance, and that the threat of reconstitution will motivate staff in low-performing schools to use more effective practices.” As did Brady and Peterson, Levy found results to be extremely variable but mostly unsuccessful.

Levy reports that one successful example, part of a court-ordered desegregation effort in San Francisco in 1984, was the result of intensive energy, long-term planning and resources (Levy, 2001): She cites Bacon’s (1997) study of the effort: “Students also had to reapply to the reconstituted schools, and the district conducted outreach efforts throughout the city. Parents of chosen students had to agree to help them get to school on time and do their homework. Most important, the state Board of Education, sued along with the district, agreed to provide extra funding to ensure the success of desegregation and the reconstituted schools.” Levy points out that a later effort in San Francisco (in 1993) had success that was “mixed, at best,” because it had a much shorter planning process, hired inexperienced teachers and failed to marshal sufficient resources to do the job.

Changing the student body seems to be the surest route to “success.” The Chicago newspaper *Substance* (2001) documented that the only reconstituted schools which showed greatly increased test scores were those at which the student bodies were replaced. The schools to which students in the failing schools were moved showed corresponding declines in test scores. *Substance* also has documented that scores actually declined in the years after restructuring at schools where the student body did not change.

An example provided by Mary Levy of a reconstitution effort with a poor outcome may be most relevant to the goals and methods of NCLB. The anonymous district she cites underwent reconstitution in the late 1990s. The reconstitution effort was plagued by problems resulting from inadequate resources and staffing issues, including excellent teachers declining to reapply because they felt they had been disrespected by the process and a new staff characterized by inexperience.

Among the conclusions drawn by researchers who followed the process: “The goals were ambitious, while fiscal, human and informa-

One successful example, part of a court-ordered desegregation effort in San Francisco in 1984, was the result of intensive energy, long-term planning and resources.

The Chicago newspaper Substance documented that the only reconstituted schools which showed greatly increased test scores were those at which the student bodies were replaced.

tional resources were insufficient, as was time. Planning time was insufficient and the three-month period from May to August was unrealistic for re-staffing and re-designing the schools. Time for professional development, teacher collaboration, and planning within the schools was drained by testing and assessment schedules, coursework for new teachers seeking certification and reporting requirements.”

Teacher Linda Christensen has written movingly about the experience of being told to pack her things after 22 years of teaching in a Portland, Oregon, school tagged for reconstitution. “The decision to leave Jefferson was not an easy one. As part of the reconstitution process, staff could reapply for their positions. In meetings packed with emotion, Jefferson staff members struggled between staying for our students who did nothing to bring this about and leaving because it was an ‘insult to reapply for a job we did well’” (Christensen, 2003). Christensen herself chose to leave her school. While she was viewed as highly qualified, as evidenced by the fact that she held leadership instructional positions for the district both before and after the reconstitution, she objected to the way her colleagues were treated in the course of the reconstitution and recognized the effort as a “quick fix” approach, “the kind now promoted by the No Child Left Behind federal legislation,” Christensen wrote. As she put it, “I packed my boxes because I refused to offer legitimacy to a shallow, mean-spirited educational policy.”

The research of Levy and others raises serious questions about the likelihood of widespread success for restructuring efforts. Assuming restructuring can be made to work without changing the student body, it clearly can only work with substantial time, resources and commitment. Under NCLB’s conditions, there is no reason to believe sufficient resources will be available.

Privatization

Consistent with reports on other NCLB remedies for failing schools, the first federal study to compare student achievement between privately managed and publicly run schools found both successes and failures. The 2003 U.S. General Accounting Office (GAO) study looked at reading and mathematics scores on standardized tests and other data to compare schools.

In meetings packed with emotion, Jefferson staff members struggled between staying for our students who did nothing to bring this about and leaving because it was an ‘insult to reapply for a job we did well’”

-Linda Christensen

Assuming restructuring can be made to work without changing the student body, it clearly can only work with substantial time, resources and commitment. Under NCLB’s conditions, there is no reason to believe sufficient resources will be available.

Marnie Shaul, director of education, workforce, and income security issues for the GAO, cautioned that the report examined a small sample, looking at just 14 privately managed elementary schools. There are a total of 417 public schools nationwide that are managed by 47 for-profit companies. She concluded, “This report shows that some privately managed companies are having better results than traditional schools, but in some cases they didn’t” (Borja, 2003).

Another study of public versus private schools deflated the view that private schools respond better than public entities to their “customers.” Researchers Luis A. Benveniste, Martin Carnoy and Richard Rothstein (2003) studied 16 public, private and charter schools that serve similar student populations in California and found little difference among them.

Their findings challenged the widely held view that private schools thrive because they have more freedom to hire and fire or to try innovative teaching methods. Parochial schools serving low-income pupils, for example, had curricula that were rigidly controlled by church leaders. Also, private school administrators were as reluctant to fire teachers as public school administrators for similar reasons: difficulty finding replacements and fear of lawsuits. “In the absence of the religious imagery on the walls, it’s really hard to tell whether you’re in a private school or a public school,” Benveniste informed *Education Week* (Viadero, 2003).

Instead, Benveniste and his colleagues (2003) found the differences among schools were tied primarily to socioeconomic differences in student bodies. Schools that served higher income pupils, whether public or private, had more in common with one another than private or public schools serving low-income pupils.

Turning schools over to private management companies means employing firms such as Edison. Edison claimed it would do a better job for less money and would be rewarded in the financial markets. Instead, failing to produce the gains it predicted and running into often massive opposition in cities such as New York and Philadelphia, Edison’s stock plummeted, with the company staying alive primarily through bailouts from wealthy supporters.

“In the absence of the religious imagery on the walls, it’s really hard to tell whether you’re in a private school or a public school.”

-Luis A.
Benveniste

Schools that served higher income pupils, whether public or private, had more in common with one another than private or public schools serving low-income pupils.

Private management is not the same as voucher programs, but the two share similar notions of private control and the argument that competition will make private schools do a better job. Education researcher Gerald Bracey (2003), writing in *Phi Delta Kappan*, offers a good summary of recent research on voucher programs, finding little conclusive evidence for voucher proponents to boast about, although he notes that does not seem to have stopped them. He quotes Jay Greene, for example, writing about evaluations of voucher programs in Charlotte, Dayton, Milwaukee, New York, and the District of Columbia that “None of them finds students harmed by receiving a voucher.” Bracey says what Greene meant by that was that test scores had not declined, a statement that is not wholly true for students in New York, Dayton and Washington, D.C. Bracey (2003) concludes, “All of this dancing around about whether vouchers work is, of course, largely irrelevant to the larger aim of voucher advocates: to privatize the public schools.”

Privatization Proponents Eager for a Boost

Despite questions about the efficacy of any interventions and a lack of capacity to monitor them, advocates of privatization remain enthusiastic about NCLB’s potential to shift resources from public to private hands. A group called Reason Public Policy Institute, which describes itself as a “public policy think tank promoting choice, competition, and a dynamic market economy as the foundation for human dignity and progress,” sees much reason for optimism in the reconstitution provisions of NCLB. Writing in the group’s publication, *Privatization Watch*, Lisa Snell (2003) predicts that NCLB will lead to more charter schools, voucher and school choice programs, and schools managed by for-profit organizations.

Snell comments on the difficulty states are having keeping up with NCLB requirements and sees a boon for the charter school movement in their predicament. “There are more than 6,000 failing schools currently listed by the federal Department of Education, and education analysts predict that this number will continue to grow as schools face tougher accountability requirements,” Snell writes. “One competitive approach would be to mandate that every state reconstitute failing schools as charter schools. The state would open a competitive bidding process to charter operators for the opportunity to create a new school in the previous school site.”

“All of this dancing around about whether vouchers work is, of course, largely irrelevant to the larger aim of voucher advocates: to privatize the public schools.”

-Gerald Bracey

Snell comments on the difficulty states are having keeping up with NCLB requirements and sees a boon for the charter school movement in their predicament.

State Takeover

As a result of a state law passed in 1987, New Jersey pioneered state takeovers of “failing” school districts. Jersey City was the first district to be taken over, in 1989, followed by Paterson in 1991 and then Newark in 1995. All were identified for takeover as a “last resort” because they were deemed unable or unwilling to correct problems identified by the state’s accountability process. In districts taken over by the state, local boards of education were removed, the positions of the chief school administrator and other executive administrative staff abolished, and a state district superintendent appointed. All three districts remain under state control.

A major study headed by Paul Tractenberg (2002) concluded that New Jersey’s interventions were “ill-conceived and poorly executed.” Takeovers in New Jersey and elsewhere “seem to have resulted in more documented successes in management and fiscal areas than in education programs.” The report notes some modest gains in “pupil performance,” but explicitly criticizes the state for overemphasizing test scores: “State test results provide only a crude measure of student performance over time, and they provide even less useful information about school district performance in areas of administration and governance. Reliance on student test scores has distracted both the state and the state-operated districts from searching out more meaningful standards for measuring progress.” Another conclusion was that interveners should focus on building the capacity of local administrators to make improvements if they do not intend long-term operation of the districts.

This, of course, requires adequate resources, say the authors. “Among other things, that means state agencies have to be well-staffed with the right kinds of personnel to assist local districts in building their capacity; collaborations with business, higher education, the nonprofit sector and community organizations have to be pursued; and local districts must have the wherewithal to employ competent staff for administrative, supervisory and support as well as instructional positions, and to provide them with strong professional development programs” (Tractenberg, *et al.*, 2002).

Massachusetts provides an example and reality check on the question of whether state governments have the capacity to intervene

A major study headed by Paul Tractenberg (2002) concluded that New Jersey’s interventions were “ill-conceived and poorly executed.”

“State test results provide only a crude measure of student performance over time, and they provide even less useful information about school district performance in areas of administration and governance.”

- Paul
Tractenberg

in the large numbers of schools and districts that will eventually be eligible for the ultimate sanctions provided for by NCLB.

Massachusetts Education Commissioner David Driscoll, lauded by federal education officials for his enthusiastic embrace of test-based reform and his support for NCLB, recently testified in a state court school finance case. The commissioner said the state education department lacks the resources to monitor and track NCLB. He said that while the state Department of Education had 1,000 employees in 1980, it has fewer than 400 today. Driscoll testified that a state panel has reviewed 46 of the 208 Massachusetts schools that have been identified so far as in need of improvement. “We can’t handle 208,” he said. “Nor do we intend to.” The state has about 1,900 public elementary, middle and high schools, according to the U.S. Department of Education (Ring, 2003).

Similarly, Maine’s Deputy Education Commissioner, Patrick Phillips, was quoted by the *Portland Press Herald* saying the state lacks anywhere near enough resources to support the numbers of schools that are likely to face intervention (Portland Press Herald, 2003). Given the fiscal constraints on school budgets around the nation, the outlook for Massachusetts and Maine is likely more the rule than the exception. The authors of a *Washington Monthly* article on states’ lack of capacity to overhaul “failing” schools put it this way: “The state agencies that NCLB relies on to carry out its sweeping mandates simply don’t have the capacity to do so. Like 220 volts of current being forced through a 110-volt kitchen appliance, the system is becoming overloaded, and the smoke is rising” (Tucker and Toch, 2004).

The fact that the law allows for restructuring of entire districts raises its own set of serious issues, including looming questions of capacity and resources. The New Jersey experience demonstrates that taking over a district is not a panacea by any stretch. Many districts like Chicago, Boston, Detroit and Baltimore that have been taken over continue to struggle to show sufficient progress. Once the districts are taken over and fail to improve, what then? Should Boston revert to having an elected school board if the district fails to make AYP over five years, as is likely to be the case? Or will politicians supportive of privatization try to hand the district over to unsuccessful entrepreneurs such as Edison?

Driscoll testified that a state panel has reviewed 46 of the 208 Massachusetts schools that have been identified so far as in need of improvement. “We can’t handle 208,” he said. “Nor do we intend to.”

“The state agencies that NCLB relies on to carry out its sweeping mandates simply don’t have the capacity to do so. Like 220 volts of current being forced through a 110-volt kitchen appliance, the system is becoming overloaded, and the smoke is rising.”

-Mark Tucker
and Thomas
Toch

Turning Around Doesn't Mean You're Going in a Better Direction

Harvard School of Education's Richard Elmore (2003), writing in *Education Leadership*, is skeptical about the prospects for drastic improvement as a result of NCLB's ultimate sanctions. "The research on how to turn around failing schools is weak, as are the state and local policies and programs designed to address this problem. If one can draw any conclusion from that research, it is that a small number of schools may emerge from classification as failing schools, that some of these will quickly return to failing status, and that only a few will continue to improve after they have emerged from failing status. Many so-called 'turnaround' schools are, in fact, functioning only at the minimal level required to keep them from returning to failing status. Turning around failing schools, in other words, is not the same as improving them."

The absence of any systematically successful means of addressing schools that are dysfunctional – not doing well and not improving, despite having adequate resources and opportunity – is a serious problem. It cannot be solved, as NCLB would attempt, by approaches that rely on ideology instead of evidence.

Whether other approaches, such as breaking existing schools into small schools, will work also needs further investigation. Preliminary evidence suggests it is harder to make this approach succeed than starting a small school from scratch. But there are at least many examples of small schools that are having substantial success with students who have not thrived elsewhere

Conclusion

Drastic "restructuring" measures appear, on balance, not to have a track record of success. For legislation that purports to support the use of "science" in research, NCLB's reconstitution requirements are clearly hope and politics, not science or education.

More fundamentally, NCLB is flawed because its vision of accountability and improvement rests almost entirely on testing and punishing. Standardized tests are wholly inadequate tools for guiding genuine school improvement. A policy based on attacking educators does not promote high-quality teaching in a challenging and supportive school.

Many so-called 'turnaround' schools are, in fact, functioning only at the minimal level required to keep them from returning to failing status. Turning around failing schools, in other words, is not the same as improving them."

-Richard Elmore

Whether other approaches, such as breaking existing schools into small schools, will work also needs further investigation.

References

- ACORN. November 21, 2002. "Parents Left Behind: A Study of State, Federal, and School District Implementation of No Child Left Behind." Available online at <http://www.acorn.org/acorn10/betterschools/BetterSchoolsReports/parents/index.php>
- ACORN. May 21, 2003. "Leaving Teachers Behind: How a Key Requirement of the No Child Left Behind Act (Putting a Highly Qualified Teacher in Every Class) Has Been Abandoned." Available online at <http://www.acorn.org/acorn10/betterschools/BetterSchoolsReports/ltb.htm>
- Advocates for Children. October 14, 2003. "Serving Those Most in Need Or Not? A Report on the Implementation of the NCLB's Supplemental Education Services in New York City."
- Alexander, K., Entwisle, D. and Olson, L. 1997. *Children, Schools & Inequality*, Boulder, CO: Westview Press.
- American Evaluation Association. November 25, 2003. Response to U.S. Department of Education, Notice of proposed priority, Federal Register RIN 1890-ZA00, November 4, 2003, "Scientifically Based Evaluation Methods."
- Anderson, M. April 28, 2004. Personal email exchange.
- Asimov, N. October 9, 2003. "Few Parents Seize Chance to Transfer Schools," *San Francisco Chronicle*. Available online at <http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2003/10/09/MN198701.DTL>
- Bacon, D. October 22, 1997. "Reconstitution – The Clint Eastwood Solution for Low-Performing Schools," *Workplace*.
- Banchemo, S. and Cholo, A. August 7, 2003. "Failing Schools Scramble to Obey Law," *Chicago Tribune*.
- Benveniste, L. A., Carnoy, M. and Rothstein, R. 2003. *All Else Equal: Are Public and Private Schools Different?* New York: RoutledgeFalmer.
- Borg, L. October 10, 2003. "Schools Decline in Annual Rankings," *Providence Journal*. Available online at http://www.projo.com/news/content/projo_20031010_rank10.69100.html
- Borja, R. R. November 12, 2003. "Report Examines Privatized Schools' Scores," *Education Week*.
- Bracey, G. January 16, 2002. "What They Did on Vacation," *The Washington Post*.
- Bracey, G. October 2003. "The 13th Bracey Report on the Condition of Public Education," *Phi Delta Kappan*.

- Brady, R. C. January 2003. "Can Failing Schools Be Fixed?" Thomas B. Fordham Foundation, Washington, D.C.
- Center on Education Policy. October 2003. "Implementing the No Child Left Behind Act." Available online at <http://www.cep-dc.org>.
- Chicago Public Schools. September 14, 2003. Press release, "67 Percent of CPS Schools – 73 Percent of NCLB Choice Schools – Score Higher on State Tests." Available online at http://www.cps.k12.il.us/AboutCPS/PressReleases/August_2003/NCLB.html.
- Christensen, L. July 2003. "Reconstituting Jefferson: Lessons on School Reform," *Rethinking School Reform: Views from the Classroom*, edited by Linda Christensen and Stan Karp.
- Clowes, G. A. September 1, 2003. "Schools Under Pressure Look to Private Sector for Help," The Heartland Institute.
- Elmore, R. F. November 2003. "A Plea for Strong Practice," *Education Leadership*, Volume 61, Number 3.
- FairTest Examiner. Fall 2002. "ESEA Controversies Continue." Available online at <http://www.fairtest.org>.
- FairTest Examiner. Winter-Spring 2003. "Transfers and Tutoring." Available online at <http://www.fairtest.org>.
- Fuller, B. et al. April 2003. "Charter Schools and Inequality: National Disparities in Funding, Teacher Quality, and Student Support," *Policy Analysis for California Education*, University of California, Berkeley and Davis, Stanford University.
- Garza, C. August 26, 2003. "Fewer Students, Money Left Behind," *Florida Times-Union*.
- Gewertz, C. February 25, 2004. "Tutoring Aid Falling Short of Mandate," *Education Week*.
- Gottlieb, R. October 5, 2003. "Promise of School Choice Falls Short," *Hartford Courant*.
- Gross, J. November 6, 2003. "Free Tutoring Is Reaching More Students in the System," *The New York Times*.
- Grossman, K. June 3, 2003. "Law to Help Kids Hurts School Scores," *Chicago Sun Times*.
- Herszenhorn, D.M. October 15, 2003. "Fewer Students are Enrolling for Tutoring," *The New York Times*.
- Herszenhorn, D. M. January 13, 2004. "Report Faults Charter Schools in New York," *The New York Times*.
- Kaplan, R. December 22, 2003. Personal email communication.

- Lashway, L. June 2003. "The Mandate to Help Low-Performing Schools," ERIC Digest 169. Available online at <http://eric.uoregon.edu/publications/digests/digest169.html>.
- LEADS. May 2003. "Reconstitution Reconsidered," Center for Education Policy and Leadership (CEPAL) at the University of Maryland.
- Levy, M. 2001. "Reconstitution: A Research Paper," for DC VOICE, Washington Lawyers' Committee for Civil Rights.
- Lutton, L.. October 9, 2003. "Schools Forced to Cut Tutoring to Buy Tutoring," *The Star*.
- MacDonald, M. August 28, 2003. "Few Apply to Transfer from 'Poor-Performing' Schools," *The Atlanta Journal-Constitution*.
- Macdonald, M. October 24, 2003. "No Child Left Behind: 264 Students Expected to Get Free Tutoring," *The Atlanta Journal-Constitution*.
- Marks, A. July 21, 2003. "As Schools 'Fail,' Parents Talk Transfers," *Christian Science Monitor*.
- Mezzacappa, D. July 7, 2003. "Law Creates Opportunities for Students, Businesses," *Philly.com*. Available online at <http://www.philly.com/mld/philly/living/education/6247079.htm>.
- Mezzacappa, D. October 26, 2003. "Vallas Critical of Tutoring Options," *Philly.com*. Available online at <http://www.philly.com/mld/inquirer/living/education/7102892.htm>.
- Miner, B. Spring 2003. "Privatizers' Trojan Horse," *Rethinking Schools*. Available online at http://www.rethinkingschools.org/archive/17_03/KPSP173.shtml.
- Miron, G. *et al.* October 2002. "Strengthening Pennsylvania's Charter School Reform: Findings from the Statewide Evaluation and Discussion of Relevant Policy Issues," The Evaluation Center, Western Michigan University, Kalamazoo, MI.
- Miron, G. and Horn, J. 1999. "Evaluation of the Public Schools Academy Initiative, Final Report," The Evaluation Center, Western Michigan University, Kalamazoo, MI.
- National Education Association. March 24, 2004. "Ten Changes Needed To Improve NCLB." Available online at <http://www.nea.org/lac/esea/tenchanges.html>.
- New York State United Teachers. February 12, 2003. "NYC, Albany Lawsuit Pushes Parental Choice," *New York Teacher*. Available online at <http://www.nysut.org/newyorkteacher/2002-2003/030212parentalchoice.html>

- New York State United Teachers. November 2003. *Not So Fast: The Unmet Challenges for Charter Schools and Those Who Oversee Them*. Available online at <http://www.nysut.org/media/releases/20031109charterschools.html>.
- North Carolina Center for Public Policy Research. 2002. "Evaluating Charter Schools in North Carolina."
- Paige, R. September 24, 2003. *Online Chat, Education Week on the Web*.
- Peterson, K. D. December 1998. "School Reconstitution: Challenges and Opportunities," *Reform Talk*, a publication of the Comprehensive Regional Assistance Center Consortium – Region VI, Issue 12. Available online at http://www.wcer.wisc.edu/ccvi/pub/ReformTalk/Year_1998/Dec_98_Reform_Talk_12.html.
- Portland Press Herald*. November 12, 2003. "State Lacks Funds to Aid Schools that Fail."
- Princeton Review Press Release. November 5, 2003. "The Princeton Review Reports Strong Q3 Results and Forecasts Stronger 2004."
- PURE (Parents United for Responsible Education). 2003. Chicago, IL. Communications to FairTest, fall.
- Rado, D., Olszewski, L. September 13, 2003. "Most in failing schools will be denied tutors," *Chicago Tribune*. Available online at <http://www.chicagotribune.com/news/chi-0309130289sep13,1,1756347.story>.
- Rasheed, A. September 20, 2003. "400 Students to Transfer from Failing Schools," *The Times-Picayune*.
- Richard, A. February 18, 2004. "Class-Size Reduction Is Slow Going in Fla.," *Education Week*.
- Ring, D. October 28, 2003. "Education Chief Takes the Stand" *The Republican*.
- Robelen, E. W. January 21, 2004. "More Students Transfer Schools, But Total Is Small, Study Reports," *Education Week*.
- Rossi, R. September 26, 2003. "Few Parents Responded to Offer of Tutoring for Children," *Chicago Sun-Times*.
- Rossi, R. October 27, 2003. "Feds: City Falling Short of School Transfer Goals," *Chicago Sun-Times*.
- Rossi, R. April 20, 2004. "Student Transfers Face Tough Odds," *Chicago Sun-Times*.
- Slovacek, S. P. March 2002. "California Charter Schools Serving Low-SES Students: An Analysis of the Academic Performance Index," California State University, Los Angeles.

- Smith, P., Molnar, A., Zahorik, J. September 2003. "Class Size Reduction in Wisconsin: A Fresh Look at the Data," Education Policy Research Unit (EPRU), Arizona State University.
- Snell, L. March 24, 2003. "The No Child Left Behind Act and the Charter School Movement," *Privatization Watch*.
- Snyder, S. October 12, 2003. "'No Child Left Behind' Law Bumps into Hard Reality," *Philadelphia Inquirer*.
- Solomon, L.C. et al. March 16, 2001. "Does Charter School Attendance Improve Test Scores?" Goldwater Institute, Phoenix, AZ.
- Substance. June 2001. "Central Office Spin Grows as Test Scores Decline."
- Sylvan employment ad. 2004. Available online at <http://www.thejobresource.com/healthcare/job.asp?job=84427000>. Accessed March 26, 2004.
- Tofig, D. August 10, 2003. "More Schools Must Offer Transfers," *The Atlanta Journal-Constitution*.
- Tractenberg, P. L. May 23, 2002. "Developing a Plan for Reestablishing Local Control in the State-Operated School Districts," Institute on Education Law & Policy, Joseph C. Cornwall Center for Metropolitan Studies, Rutgers-Newark. Available online at <http://www.state.nj.us/njded/schools/sosd/local.htm>.
- Tucker, M., Toch, T. March 2004. "Hire Ed," *Washington Monthly*.
- Viadero, D. January 22, 2003. "Public or Private, Study Finds Schools Similar," *Education Week*.
- Walsh, M. January 23, 2002. "Tutoring Services See Opportunity in New Law," *Education Week*. Available online at <http://www.edweek.org/ew/newstory.cfm?slug=19tutor.h21>.
- Walsh, M. July 9, 2003. "Judge Rejects Suit Seeking Transfers under Federal Law," *Education Week*.
- Winerip, M. October 1, 2003. "In 'No Child Left Behind,' a Problem With the Math," *The New York Times*.

IV. Money, Education and Accountability

A widespread critique of NCLB is that its promise of improvement is meaningless given the insufficient resources allocated to the education of low-income children. In addition to continuing inequitable and inadequate school funding, many students suffer from the effects of poverty that schools alone cannot solve. As Susan Neuman, President Bush's former assistant secretary for elementary and secondary education, put it, "When we say all children can achieve and then not give them the additional resources [needed for meeting that goal] we are creating a fantasy" (Manzo, 2004).

"When we say all children can achieve and then not give them the additional resources [needed for meeting that goal] we are creating a fantasy"

- Susan Neuman

A. Funding NCLB and Funding Education

While President George W. Bush claims to be the "education president," touting NCLB as his credential, there is widespread agreement that NCLB's funds fall far short of what would be needed to make every child "proficient." The Fiscal Year 2003 federal Title I appropriation was \$11.3 billion. The Bush administration sought an increase to \$12.3 billion for FY 2004, far less than the \$18 billion Congress authorized when it created NCLB. In March 2004, the National Conference of State Legislatures calculated the gap between funds authorized and appropriated for NCLB at \$9.6 billion (Mandate Monitor, 2004). Bush claimed the \$1 billion increase he sought for FY 2004 was "more than enough money" to accomplish NCLB's goals.

However, according to William J. Mathis (2004), superintendent of schools in Brandon, Vermont, and a professor of education finance at the University of Vermont, "The law says that each child living in poverty is eligible to receive an extra 40 percent of the state's average per-pupil spending. The Congressional Research Service calculates this would amount to \$30.4 billion. Thus, the law is funded at only 41 percent of its own definition." An increase of \$30.4 billion would be welcome, but it is not likely to be enough to bring all children to the proficient level.

In March 2004, the National Conference of State Legislatures calculated the gap between funds authorized and appropriated for NCLB at \$9.6 billion

Researchers have employed a variety of methods to estimate what it would take to get every child to pass mandated state tests. According to Mathis, a conservative estimate is that a 20 percent increase in spending nationwide, an increase of about \$84.5 billion, would be needed (Mathis, 2003). Elsewhere, Mathis (2004) points out that studies in 13 states that have considered the costs of bringing all children

up to proficient project “new money increases” of 20 to 40 percent, with a median increase of 30.2 percent. Nationally, that would require an increase of more than \$120 billion.

Many state policymakers charge that the modest federal budget increase will not cover the costs of fulfilling the law’s mandates. The most detailed study to date on the costs of NCLB, a report by the Ohio Department of Education, concluded it would cost the state an additional \$1.5 billion per year to administer NCLB and meet its achievement goals – twice what the state now gets from the federal government (Hoff, 2004). Mark Joyce, executive director of the New Hampshire School Administrators Association, estimated that his state’s schools will receive an average of \$77 per student in new federal money, but complying with the law’s mandates will cost \$575 per student (New Hampshire School Administrators Association, 2002).

A proposed lawsuit drafted by the National Education Association in July 2003 points out that the failure to allocate enough funds to pay for mandated testing, tutoring and transfers means the federal government is violating the terms of the NCLB law, which specifies in Sec. 9527, “Nothing in this Act shall be construed to authorize an officer or employee of the Federal Government to...mandate a State or any subdivision thereof to spend any funds or incur any costs not paid for under this Act.” To date, however, no state has filed suit on these grounds.

Some proponents of the NCLB test-and-punish philosophy espouse the idea that the nation should stop throwing money at the failed public education approaches of the past and hold everyone to the same high standards once and for all (Paige, 2003). Texas columnist Molly Ivins describes that viewpoint as: “Spend less money, and bring back the lash and McGuffey Readers” (Ivins, 2003).

Some of the same people who have argued the federal government doesn’t need to spend more on education are also fighting against spending increases at the state and local level. For example, Republican former House Majority Leader Dick Armey leads the group Citizens for a Sound Economy, which fights tax increases for education. The group helped kill a \$1.2 billion tax hike in Alabama, earmarked to improve public education, and it helped derail an Oregon tax increase that lawmakers intended to provide 13 percent more money for public schools (Straight A’s, 2003).

A proposed lawsuit drafted by the National Education Association in July 2003 points out that the failure to allocate enough funds to pay for mandated testing, tutoring and transfers means the federal government is violating the terms of the NCLB law.

Some of the same people who have argued the federal government doesn’t need to spend more on education are also fighting against spending increases at the state and local level.

Federal funds have never covered more than a small fraction of the nation's total elementary and secondary education spending. Currently, the federal share is 7 to 8 percent. Some NCLB proponents argued the law would force states to live up to their obligations to adequately fund public education. According to *Education Week's Quality Counts* (2004, 2003) most states exhibit sizeable funding differences among their districts, with poorer districts having lower-funded schools. However, during the first two years of NCLB, most states cut education funding due to their fiscal crises.

A report by the Vermont Society for the Study of Education concludes that the consensus among school funding experts is that students in poverty require twice the current per pupil expenditure to attain proficiency (Mathis, 2002). No state has come close to providing such support for children from low-income families.

Secretary of Education Rod Paige has correctly and repeatedly charged this country with perpetuating educational apartheid. However, Paige defends a law that not only fails to redress the funding apartheid that produces unequal outcomes, but may in fact exacerbate such inequities.

An *Oakland Tribune* article (Tucker, 2003) vividly described what educational apartheid looks like in California: "Students in poor communities enter dilapidated classrooms where uncredentialed teachers with inadequate materials await — and where parent involvement is limited or nonexistent. In better-off neighborhoods, sometimes just a few miles away, the schools nearly sparkle, sporting the latest facility upgrades, top-notch equipment and the most experienced teachers. With nighttime PTA meetings, weekend potluck fund-raisers and various festivities, these better schools lure upwardly mobile home buyers drawn to the first-rate education and other opportunities offered to their kids."

Funding disparities are being legally challenged in a number of states. New York's highest court recently ruled that the state has failed to provide adequate funding to New York City and other urban, poor districts. The court requested that a study be conducted to determine the cost of an adequate education in the state. In March 2004 (American Institutes for Research, 2004), researchers released a preliminary report calculating the state would need to

The consensus among school funding experts is that students in poverty require twice the current per pupil expenditure to attain proficiency.

Paige defends a law that not only fails to redress the funding apartheid that produces unequal outcomes, but may in fact exacerbate such inequities.

spend an additional \$6.21 billion (an increase of 19.4 percent), including an additional \$3.62 billion in New York City alone.

In California, a major educational adequacy and equity case, *Williams*, is before the courts. New Hampshire is engaged in a long-running battle over how to fund a decision by the state's top court requiring additional funds for low-income districts to ensure that all students receive "an adequate education." In Massachusetts, a Superior Court judge has ruled that the state is not providing enough funding to adequately educate students in 19 low-income districts and recommended that the state's highest court hear a lawsuit filed by these districts against the state Department of Education (Maguire, 2004). A group of districts in Kentucky recently filed suit, charging that funding gaps between districts have widened since that court found the state's education system to be inadequate. And the Ohio Supreme Court reaffirmed its decision mandating changes in school funding — but once again did not mandate any specific aid level.

A Kansas trial court ruled in December 2003 that the state education finance system is unconstitutional due to its failure to "provide equity in funding for all Kansas children." Although a court had deemed the funding system constitutional in 1994, legislative changes in the intervening years led to funding disparities exceeding 300 percent. Hoover Institution fellow Dr. Eric Hanushek, a noted proponent of the idea that money doesn't matter in education, testified for the state. However, he was forced to acknowledge in his testimony that "money spent wisely, logically, and with accountability would be very useful indeed." He concluded by agreeing with the statement: "Only a fool would say money doesn't matter" (School Funding Updates, 2003).

And in a report whose title – *Can Separate Be Equal?* – is an implicit rebuttal to Rod Paige's use of the issue of apartheid, the Century Foundation argued, "Fifty years after *Brown v. Board of Education*, NCLB is an effort, like most education reform, to make separate but equal work."

B. Schooling and Poverty

While states continue to allow great inequality in school funding, the social conditions that make education in low-income communities a more challenging task are also worsening. Unemployment is up, states are cutting access to medical care and, like the federal govern-

"Only a fool would say money doesn't matter"

-Dr. Eric Hanushek

"Fifty years after Brown v. Board of Education, NCLB is an effort, like most education reform, to make separate but equal work."

– *Can Separate Be Equal?*

ment, cutting housing support. A recent *New York Times* story reported that in some states, there has been a 50 percent increase in homeless students in the past year. Schools are increasingly faced with trying to compensate for these students' unmet basic needs (Dillon, 2003). Schools thus face major mandates to improve educational outcomes - or at least test scores - for low-income and minority-group children even as their funding declines and other social problems that affect learning intensify.

Failing to address the fundamental problems that undergird school inequality enables policymakers to blame teachers for the consequences of deep-seated social ills. It enables the federal government to pass legislation named "No Child Left Behind" without addressing housing, nutrition, health care or other basic needs that are unmet for millions of girls and boys.

Going at least as far back as the famous Coleman Report of 1966, research has consistently found a very high correlation between socioeconomic status and educational outcomes, including test scores. Analyzing the same data, Christopher Jencks and his colleagues (1972) concluded that schools were not good vehicles for addressing social inequality. Richard Rothstein has regularly argued that poverty is a more significant cause of low academic achievement than are weak schools, and that providing low-income students with things like dental care may do more for academic achievement than test preparation (Rothstein, 2001). George Schmidt, publisher of the Chicago monthly education paper *Substance*, reminds us that the language of "achievement gap" masks the real issues of poverty and racism that cause the gap (Schmidt, 2003).

Mathis (2003), writing in *Phi Delta Kappan*, offers evidence for this view, noting that the United States ranks 21st out of 24 industrialized nations in educational equality. He refutes the argument that schools alone can close the achievement gap: "Simply teaching children will have little effect if they return to bad neighborhoods, single-parent homes, foster care, inadequate health care and a general lack of support." Mathis referenced research by Whitney Allgood and Richard Rothstein demonstrating that overcoming the effects of poverty requires broad-scale interventions, including community clinics, before and after school programs, early childhood intervention and summer school programs.

In some states, there has been a 50 percent increase in homeless students in the past year. Schools are increasingly faced with trying to compensate for these students' unmet basic needs

Failing to address the fundamental problems that undergird school inequality enables policymakers to blame teachers for the consequences of deep-seated social ills.

Valerie Lee and David Burkam (2002) found that the academic achievement gap is largely in place before children enter kindergarten and first grade. They identified a series of social factors that cause race and class differences in school performance.

Educational Testing Service researcher Paul Barton (2003) identified a series of school and non-school factors that substantially contribute to the test score gap. The out-of-school factors include lead paint, which is much more likely to be found in housing occupied by low-income families, and high student mobility, which is often caused by low-income parents having difficulty making rent payments.

A recent RAND Corporation report explained that approaches like NCLB, which address school factors alone, are unlikely to combat the root causes of the achievement gap: “NCLB assumes that educators can address the factors that have contributed to low levels of achievement effectively and that the way to make this happen is through a combination of high standards, powerful incentives, and technical assistance. If this is not the case — for example if poor achievement is due to high levels of student mobility — this model of accountability is not likely to be an effective strategy for improvement. Neither incentives nor the options for supplemental services or parental choice are likely to address the root cause of the problem” (Stecher et al, 2003).

Despite these well-supported arguments about socioeconomic influences on academic achievement, the U.S. has consistently sought to address social inequality through education. Thus, elementary and secondary schools have become the main arenas in which social differences of class and race are supposed to be addressed. The Elementary and Secondary Education Act (ESEA), which passed under President Lyndon Johnson, was one such effort. Decades of evidence shows that while ESEA may have prevented gaps from increasing, achievement differences as measured by standardized tests such as the National Assessment of Educational Progress have not narrowed significantly.

Over the years, various individuals and organizations, most recently groups such as the conservative Heritage Foundation and the more centrist Education Trust, have argued that schools can do a far better job educating students from low-income backgrounds with the resources they have. The evidence to support their claims is contested, with some researchers responding that the numbers of “successful”

“If poor achievement is due to high levels of student mobility — this model of accountability is not likely to be an effective strategy for improvement. Neither incentives nor the options for supplemental services or parental choice are likely to address the root cause of the problem.”

*-Brian Stecher,
et.al.*

Decades of evidence shows that while ESEA may have prevented gaps from increasing, achievement differences as measured by standardized tests such as the National Assessment of Educational Progress have not narrowed significantly.

schools claimed by these groups have been greatly inflated if not largely fabricated. For example, Steven Krashen (2002) found that Education Trust (Ali and Jerald, 2001) used very low standards for identifying California schools making exemplary progress, allowing schools with test score gains in only one subject in one grade to count as significant school-level improvement. In addition, the identified schools tended to have fewer students from poor households than the average California school (Krashen, 2002).

To acknowledge these facts is not to argue that schools cannot make a substantial difference in the lives of low-income students. There are some schools, such as those associated with educators such as Deborah Meier and Ann Cook (e.g., the New York Performance Standards Consortium schools), that over the years have produced substantially better outcomes for low-income children as indicated by graduation rates, college enrollment and undergraduate success (Institute for Education and Social Policy, 2001). Meier (2002) does not claim such schools can overcome the effects of poverty to ensure equality of outcomes, but says they have made a difference in measures more significant than test scores.

The question of the extent to which schools can overcome the consequences of poverty and racism has been hotly debated at least since Ron Edmonds (1979) argued that some schools serving low-income children had test scores similar to students in more affluent schools. While evidence does suggest school systems can do better than they often have, there are no large-scale examples in which schools have been able to conquer the effects of poverty.

Funding, of course, cannot be considered apart from what the money is spent on, as those who deny the importance of persistent funding inequities remind us. In essence, the federal law requires funds to be targeted toward activities that raise test scores. But as research by Audrey Amrein and David Berliner (2002) shows, there is no real evidence that boosting scores on one test means real learning has improved. Nor does a focus on tests lead to closing the racial score gap (Johnson *et al.*, 2001).

There are some schools, such as those associated with educators such as Deborah Meier and Ann Cook, that over the years have produced substantially better outcomes for low-income children as indicated by graduation rates, college enrollment and undergraduate success.

While evidence does suggest school systems can do better than they often have, there are no large-scale examples in which schools have been able to conquer the effects of poverty.

Thus, the nation finds itself in an apparent bind: The federal government now requires states to improve results, but is providing very little money. States, which historically have allowed funding and educational inequalities to persist, have cut classroom aid. Even if schools can do a better job than ever before, there is no solid evidence that inequities can be overcome in the classroom alone, certainly not by a focus on test preparation. If states and the federal government do not jointly address social and educational inequities, districts and schools will inevitably fail to meet the new mandates.

What then? Will the U.S. grapple with the poverty that underlies low educational performance, recognize that it will take substantial funds to improve schools, and understand that test prep is not high quality schooling? Such considerations would entail vastly different national priorities. It is perhaps instructive that the Children's Defense Fund (2003), whose slogan "Leave No Child Behind" was appropriated (and put in passive tense) for the law, has crafted legislation to start to address these multiple issues — but the legislation has not moved in Washington.

Even if schools can do a better job than ever before, there is no solid evidence that inequities can be overcome in the classroom alone, certainly not by a focus on test preparation.

Will the U.S. grapple with the poverty that underlies low educational performance, recognize that it will take substantial funds to improve schools, and understand that test prep is not high quality schooling?

References

- Ali, R. and Jerald, C. 2001. "Dispelling the Myth in California." Available online at http://www2.edtrust.org/NR/rdonlyres/424BBCDB-822F-4631-B1F4-36BDFB52B956/0/DTM_Report_CA.pdf.
- American Institutes for Research. February 3, 2004. "The New York Adequacy Study," Management Analysis and Planning, Inc. Available online at www.cfequity.org/Costingout/FINALVERSION-PRELIMINARYREPORT-2004-02-03.pdf.
- Amrein, A. L., Berliner, D. C. December 2002. "The Impact of High-Stakes Tests on Student Academic Performance," Arizona State University. Available online at http://www.asu.edu/educ/eps/EPRU/epu_Research_Writing.htm
- Barton, P. October 2003. "Parsing the Achievement Gap: Baselines for Tracking Progress," Educational Testing Service, Policy Information Center. Available online at www.ets.org/research/pic/parsing.pdf.
- Children's Defense Fund. 2003. The Act to Leave No Child Behind. Available online at <http://www.cdfactioncouncil.org/act.htm>.
- Dillon, S. November 27, 2003. "School Is Haven When Children Have No Home," *The New York Times*.
- Edmonds, R. 1979. "Effective Schools for the Urban Poor," *Educational Leadership* 37(1), 15-27.
- Education Week*. 2004, 2003. *Quality Counts*. Available online at www.edweek.com. See tables on Resources: Adequacy.
- Hoff, D. J. February 4, 2004. "Debate Grows on True Costs Of School Law," *Education Week*. Available online at <http://www.edweek.com/ew/ewstory.cfm?slug=21NCLBCost.h23>.
- Institute for Education and Social Policy. December 2001. "Final Report of the Evaluation for New York Networks for School Renewal," New York. Available online at <http://www.nyu.edu/iesp/publications/nynsr/finalreport.PDF>.
- Ivins, M. July 18, 2003. "Fixing Education." *Alternet*. Available online at <http://www.alternet.org>.
- Jencks, C. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*, New York: Basic Books.
- Johnson, T., Boyden, J. E., Pittz, W. J. October 30, 2001. "Racial Profiling and Punishment in U.S. Public Schools: How Zero Tolerance Policies and High Stakes Testing Subvert Academic Excellence and Racial Equity," Applied Research Center, Oakland, CA.

- Krashen, S. February 2002. "Poverty has a Powerful Impact on Educational Attainment, or, Don't Trust Ed Trust," *Substance* (5132 W. Berteau Ave., Chicago, IL 60641; csubstance@aol.com); and available, with permission, at <http://www.fairtest.org/k12/krashen%20report.html>
- Lee, V. and Burkam, D. T. September 2002. *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, D.C.: Economic Policy Institute.
- Manzo, K. K. May 12, 2004. "Reading Experts Offer Insights into State, Federal Policies," *Education Week*.
- New Hampshire School Administrators Association Executive Board. November 19, 2002. "Analysis of Cost Impact of ESEA - No Child Left Behind Act on New Hampshire." Available online at <http://www.nhsaa.org/pdf/files/ESEAcostimpactan.pdf>.
- Maguire, K. April 26, 2004. "Judge Rules School Funding Formula Insufficient," *Boston.com*.
- Mandate Monitor. March 10, 2004. National Conference of State Legislatures. Available online at Mandate Monitor link at <http://www.ncsl.org/programs/press/2004/pr040310.htm>.
- Mathis, W. J. October 22, 2002. "The Federal 'No Child Left Behind' Law: Should Vermont Take the Money?" Vermont Society for the Study of Education: Policy Discussion Paper.
- Mathis, W. J. May 2003. "No Child Left Behind: Costs and Benefits," *Phi Delta Kappan*. Available online at <http://www.pdkintl.org/kappan/k0305mat.htm>.
- Mathis, W. J. April 21, 2004. "Two Very Different Questions: The Cost of Leaving No Child Behind? Or, the Cost of Implementing 'No Child Left Behind'?" *Education Week*.
- Meier, D., 2002. *In Schools We Trust: Creating Communities of Learning in an Era of Testing and Standardization*. Boston: Beacon Press.
- Paige, R. October 30, 2003. "It's Not About the Money," *The Wall Street Journal*.
- Rothstein, R. March 7, 2001. "Seeing Achievement Gains by an Attack on Poverty," *The New York Times*. Available online at http://www.epinet.org/content.cfm/webfeat_lessons20010307
- Schmidt, G. July 26, 2003. Post to ARN-L listserv. Available online at <http://www.intersersity.org/lists/arn-l/archives/jul2003/msg00395.html>.
- "School Funding Updates." Campaign for Fiscal Equity, December 15, 2003. Available online at <http://www.schoolfunding.info/>

“SEC. 9527. PROHIBITIONS ON FEDERAL GOVERNMENT AND USE OF FEDERAL FUNDS. “(a) GENERAL PROHIBITION.— Nothing in this Act shall be construed to authorize an officer or employee of the Federal Government to mandate, direct, or control a State, local educational agency, or school’s curriculum, program of instruction, or allocation of State or local resources, or mandate a State or any subdivision thereof to spend any funds or incur any costs not paid for under this Act. 20 USC 7907. Cite title number, code abbreviation (U.S.C.) Section number 9527, year on volume spine.

Stecher, B., Hamilton, L., Gonzalez, G. November 2003. “Working Smarter to Leave No Child Behind: Practical Insights for School Leaders,” RAND Corp., Santa Monica, CA.

“Straight A’s: Public Education Policy and Progress.” December 10, 2003. Alliance for Excellent Education, “Arkansas Takes a Look at Adequacy.”

Tucker, J. June 16, 2003. “State’s Spending Fails to Keep Pace,” *Oakland Tribune*.

“The requirements of the law will have to be changed this year either through regulation or through legislation. There is too much dissatisfaction.”
— Jack Jennings, Center on Education Policy

V. Organizing for Change Amid Growing Opposition to NCLB

The *New York Times* recently reported on just how tough it has been to sell NCLB to many Americans. It quoted Russel Sias, a retired engineer and registered Utah Republican whose daughter is a middle school teacher: “I feel like we’re hearing the best vacuum cleaner salesman in the world. They’re going to label every school in the country as failing, and they call it empowerment?” (Dillon, 2004). For NCLB to be remade into a law that will really empower parents, students and teachers to sustain and improve public education, many more citizens must learn enough about NCLB to understand that the reality does not match the rhetoric. Then they must act together to bring about change.

NCLB clearly damages public education. But it offers an opportunity as well. The opportunity is that NCLB will energize those who believe public schools are a pillar of a vibrant democracy and who know that most schools, while they may need improvement, do not deserve labels and sanctions. Those like Mr. Sias, who see NCLB undermining rather than supporting public schools, must come together and demand changes to make the law live up to its title.

For that to happen, groups of parents, educators, civil rights advocates, and other community members must work together to do three things: sharpen and popularize a critique of the law, develop a model for a new version of the law, and build a grassroots campaign to persuade Congress to overhaul NCLB.

This report is one tool for developing a critique that will mobilize citizens from many walks of life who care about improving public education. Chapter VI, on alternatives to NCLB’s punitive approach, is designed as a template for the necessary work of defining what a revised law should include. The challenge ahead is to bring together the various constituencies and help them use the critique and the model to bring about change. To do so also requires understanding the strengths and pitfalls of the emerging opposition to NCLB.

“I feel like we’re hearing the best vacuum cleaner salesman in the world. They’re going to label every school in the country as failing, and they call it empowerment?”

—Russel Sias

Groups of parents, educators, civil rights advocates, and other community members must work together to do three things: sharpen and popularize a critique of the law, develop a model for a new version of the law, and build a grassroots campaign to persuade Congress to overhaul NCLB.

A. The Emerging Movement

Media attention has focused on criticism and resistance from state legislatures. By mid-March of 2004, at least 21 states had passed or proposed measures to opt out of NCLB or to seek changes in the law (*School Board News*, March 2004). By mid-April, 27 states had seen bills or resolutions calling for changes in the law, full funding, studies of the costs, prohibiting state funding on the law, or for opting out altogether (NEA, 2004).

For example, in March, the Oklahoma House of Representatives unanimously passed a resolution calling for repeal of NCLB. An overwhelming vote to prohibit spending state funds to comply with NCLB mandates by the conservative Utah House of Representatives showed the bipartisan nature of opposition. Maine legislators followed suit, also refusing to spend state funds on NCLB. The Republican-dominated Virginia legislature voted 98 to 1 for a resolution objecting to many aspects of NCLB. In response to a growing chorus of criticism, the National Conference of State Legislatures (2004) has convened a task force to study the consequences of NCLB.

Perhaps the most audible complaints have centered on claims that the federal law is underfunded and overly intrusive. Criticism of massive federal intervention in state and local educational policy takes several forms. It includes resistance to several things: to NCLB's bureaucratic requirements; to having to alter state accountability programs to dovetail with the federal requirements; to expanding use of standardized tests; to the arbitrary Adequate Yearly Progress formula; and to the sanctions imposed for failure to make AYP.

While states have legitimate complaints about underfunding and the law's intrusiveness, it would be a mistake simply to demand that the federal government leave states to their own devices, particularly if that means continued neglect of needy districts and students. Such an approach allows the law's defenders to paint its opponents as either simply against accountability or as defenders of a form of "state's rights" that echoes the anti-desegregation battles of the 1950s, '60s and '70s. To support genuine school reform, advocates must demand adequate and equitable education for all students and not confuse arbitrary intrusiveness with proper concern for equity, civil rights, and educational quality.

By mid-March of 2004, at least 21 states had passed or proposed measures to opt out of NCLB or to seek changes in the law

To support genuine school reform, advocates must demand adequate and equitable education for all students and not confuse arbitrary intrusiveness with proper concern for equity, civil rights, and educational quality.

NCLB's harmful impact on teaching and learning has been documented in the media and many reports. But perhaps because NCLB exacerbates rather than creates the problem of standardized test overuse and misuse, this aspect has not been as prominently reported (on state test use, see Neill, 1997). Those closest to classrooms may be very concerned over the narrowing and dumbing down of education. Policymakers and their supporters in the corporate world, however, often focus on accountability by numbers rather than on education itself. In addition, education issues are complex and not readily reduced to sound bites.

In general, the media have grown more skeptical of NCLB. As with legislators, the initial focus was on inadequate funding and federal intrusiveness. It is increasingly common for newspaper editorials to remark that NCLB is a "one size fits all" approach that will not work.

Some examples from among dozens:

- "President Bush's so-called 'No Child Left Behind' Act is proving to be full of snares, contradicting state education goals, confusing and demoralizing teachers and principals, penalizing the neediest and, a new study shows, sabotaging schools with diverse student populations." – *Palm Beach Post* (2004)
- "The obvious conclusion is that a one-size-fits-all set of standards is both unrealistic and unfair. Schools need time and flexibility to bring struggling students up to their potential. They should not be forced to match a timetable set in Washington." – *Honolulu Advertiser* (2004)
- "It's time for New Mexico to leave President Bush's sad excuse for education reform behind. Despite claims to the contrary, the No Child Left Behind Act is a catastrophe in the making." – *Albuquerque Tribune* (2004)
- "It does not seem fair or reasonable to think that all students are the same or that all students can attain the same levels of proficiency. Addressing the needs of the individuals and ensuring that everyone receives a quality education – not scoring 'X' on any one test – should be the focus." – *Lansdale (PA) Reporter* (2004)

Policymakers and their supporters in the corporate world, however, often focus on accountability by numbers rather than on education itself.

Schools need time and flexibility to bring struggling students up to their potential. They should not be forced to match a timetable set in Washington.

In addition to the most visible criticisms from legislatures or editorial writers (see Appendix 1: What's in a Name?), it is clear that a groundswell of criticism is rising among teachers, principals, and many parents and other community members. When members of Congress visit their districts, many get an earful of harsh words about NCLB. This opposition has influenced organizations such as the American Association of School Administrators and the National Education Association, and is likely to have a large impact on membership groups.

B. Efforts to Fix Some NCLB Problems

Education organizations have suggested many changes to the law, most aimed at repairing specific flaws. These changes do not address the larger structural problems in the law, but if adopted they would lessen the immediate damage. The Department of Education (DOE) has already incorporated a few proposals into its regulations, such as allowing limited English proficient (LEP) students who attain proficiency to be counted in the LEP category for two more years.

The Council of Chief State School Officers (CCSSO, 2004), an organization of state school superintendents, has offered several proposals to change how NCLB identifies schools for improvement. Some have fairly specific objectives, such as classifying schools as “in need of improvement” (INOI) only when the same subgroup of students fails to meet AYP in the same subject for two or more consecutive years. Others are more general, such as permitting state accountability systems to use “multiple measures that can compensate for each other (including AYP data) in making accountability decisions, so long as the state accountability system is shown to pursue the same goal of 100 percent proficiency by 2013-2014.” The compensatory aspects of this proposal may be rendered moot by its explicit acceptance of the unrealistic goal of 100 percent proficiency. But the underlying approach opens the door to broader conceptions of accountability.

The National Education Association (NEA, 2004) has released “Ten Changes Needed to Improve ‘No Child Left Behind.’” The NEA concurs with the CCSSO on identifying schools as INOI only if the same subgroup repeatedly fails to meet AYP. The NEA’s proposals would also allow states to use multiple measures, including such things as attendance, graduation and other state and local assessments, to

A groundswell of criticism is rising among teachers, principals, and many parents and other community members.

The NEA’s proposals would also allow states to use multiple measures, including such things as attendance, graduation and other state and local assessments.

create a progress index. They would also ensure that only students in those groups that fail to make AYP would be eligible for the school transfer and tutoring provisions.

The National School Boards Association (NSBA, 2004) also calls for applying sanctions only when students in a given subgroup fail to make AYP in one subject for two or more years. NSBA would allow states to use “gain scores” rather than consider only the mandated cutoff scores. (Gain scores track and recognize individual student progress toward proficiency rather than only award “credit” when the student reaches that level.)

The National Conference of State Legislatures (NCSL) has endorsed 16 specific suggestions for change (NCSL, 2003). For example, NCSL echoed some other groups that called for allowing different starting points for AYP calculations for schools and subgroups, rather than allow only a uniform state starting point. NCSL has now established a task force to consider more than a dozen elements of the law and make further recommendations for changes (NCSL, 2004). Minnesota Senator Steve Kelley, a task force co-chair, explained that recent changes in regulations have not been sufficient: “But the fundamental problem remains with the language of the law. I hope this task force can come up with recommendations for amending and improving the way No Child Left Behind works so it really supports the school improvement efforts states are already making.”

Many of the groups support allowing states to create alternative assessments for students who are not severely disabled but who are significantly impaired (called “gap kids” in some descriptions), and not to expect those students to reach the same proficiency as their peers.

For the most part, these suggestions do not challenge the test-and-punish structure of the law. But many would make the process of implementation less onerous and reduce the damage caused by extreme rigidities in the law and regulations. Some members of these groups fear ameliorative changes would lessen the push for more substantive changes. However, the Department of Education has resisted everything that exceeds very minor changes, and no legislative proposals are expected to progress during 2004.

The National Conference of State Legislatures echoed some other groups that called for allowing different starting points for AYP calculations for schools and subgroups, rather than allow only a uniform state starting point.

Many of the groups support allowing states to create alternative assessments for students who are not severely disabled but who are significantly impaired.

C. Public Opinion

Polls and other measures of public opinion show that once people are exposed to details about NCLB, support for its general principles turns to extreme concern and opposition to many of its particulars.

For example, a national poll sponsored by Results for America found that while parents supported the overall concept of NCLB, their support evaporated when they considered what the law might mean specifically for their child's school. Nearly three out of four parents said they oppose cutting federal funds to their children's school if it were deemed to be failing, and only 13 percent favor linking federal funds to performance. The poll found these responses cut across demographics and party affiliation (Results for America, 2004).

A more recent poll by the Public Education Network and *Education Week* confirmed these findings and showed growing voter opposition to NCLB. While the number of voters who said they had heard of the law grew from 56 percent last year to 75 percent this year, the percentage who say they oppose the law increased by 20 points (Public Education Network, 2004). Celinda Lake, president of Lake Snell Perry & Associates, which conducted the survey, explained, "The more you know about NCLB, the less likely you are to be in favor of it" (*School Board News*, April 2004).

As opposition to NCLB grows, the specifics of alternative proposals matter greatly to those building an assessment reform movement. NCLB earned the support of some civil rights leaders precisely because it appeared to address the U.S. system of educational apartheid, as Secretary Paige has called it. There are real differences among NCLB opponents as to what constitutes an appropriate use of standardized testing in school accountability and what constitutes effective and helpful accountability.

Education, civil rights, and parent groups represent those who are most affected by NCLB and who have the most to gain from changing the law. A reform alliance will have to work to resolve differences by initiating discussions with not just the national leaders of these groups, but with classroom teachers, parents, and community activists. While NCLB will worsen educational apartheid, efforts to overhaul it must

A national poll sponsored by Results for America found that while parents supported the overall concept of NCLB, their support evaporated when they considered what the law might mean specifically for their child's school.

While the number of voters who said they had heard of the law grew from 56 percent last year to 75 percent this year, the percentage who say they oppose the law increased by 20 points.

promote a positive alternative, rather than simply rejecting its bureaucratic intrusiveness, over-reliance on testing, rigid AYP scheme, and underfunding.

D. Educating for Change

A grassroots movement of educators, parents and civil rights groups must collaborate on three fronts. First, such a movement can help articulate the law's harsh effects on classrooms, teachers and students. Second, it can demand that policymakers put the focus on improving teaching and learning rather than on testing and punishing.

Third, it can contribute ideas to the development of systems that hold public schools accountable to parents and students rather than to distant government bureaucrats. In other words, it can help to create a small-d democratic accountability system.

In addition to growing criticism in the media and in public opinion, there are signs of the beginnings of grassroots awareness and mobilization. In one of the first actions taken at a community level, the Nashville School Board unanimously adopted a resolution stating its "serious concerns" about some provisions of the federal "No Child Left Behind" legislation and suggested seven potential changes. Among the things the board objected to was the requirement that "all" children meet the same standard, which the board said "is unrealistic for children who do not speak English and for some special education students." The Nashville board also said the law has "the potential for unintended discrimination against some groups of children," the ones whose test scores mark their school for failure and punitive sanctions (Metropolitan Nashville Public Schools, 2003).

Massachusetts provides several effective models for grassroots action. In Brookline, an affluent yet diverse community bordering Boston, the school committee became alarmed at the details of NCLB and adopted a resolution expressing its concerns. The resolution strongly criticizes NCLB for using standardized test scores to judge students' academic performances. "While student test data and other indices of accomplishment help inform teachers and school administrators," the resolution reads, "the ESEA's system of accountability is disastrous policy that further entrenches the reliance on a single test score and imposes very problematic and costly consequences for school districts" (Brookline School Committee, 2004).

The Nashville board also said the law has "the potential for unintended discrimination against some groups of children," the ones whose test scores mark their school for failure and punitive sanctions.

"The ESEA's system of accountability is disastrous policy that further entrenches the reliance on a single test score and imposes very problematic and costly consequences for school districts."

-Brookline, MA,
School
Committee

Brookline School Committee members helped introduce a resolution critical of NCLB at the 2003 delegate assembly of the Massachusetts Association of School Committees (MASC). The resolution passed overwhelmingly. Indeed, one of the few critical comments raised during the debate on the resolution was from a delegate concerned that it was not a strong enough condemnation of the law. The resolution called for “changes in the accountability provisions of the ESEA, including amendments which move away from the annual testing requirements and severe sanctions now included in the law.” It also called on the National School Boards Association to adopt a resolution similar to the one passed by the MASC (Massachusetts Association of School Committees, 2003).

Soon after the school board’s actions, a Brookline parent group that had formed four years earlier to protest the state’s high-stakes high school exit exam held a public forum on NCLB. The Brookline Coalition for Authentic Reform in Education worked with the local teacher’s union and a number of other community groups to plan the forum. Brookline has a long tradition of strong support for and pride in its very successful public schools. The town has small class sizes, high graduation and college attendance rates, and strong test scores. Many residents had been shocked to learn that six of Brookline’s nine schools had failed to make AYP for the previous year. In some cases, this was because of their failure to meet the 95 percent participation target, and in others because limited English proficient students failed to make AYP.

In part fueled by questions and concerns about the district’s NCLB report card, and due to extensive publicity from CARE, the Brookline NCLB forum attracted close to 200 people, including representatives from both U.S. senators’ offices. With Senator John Kerry the presumptive Democratic presidential nominee and Senator Edward Kennedy a key architect of the law, the significance of their attendance was not lost on the participants. Organizers of the event also circulated a petition.

E. Organizing for Change

Effective change will require more vigorous public education coupled with grassroots resistance that is in turn linked to the growing network of organizations seeking fundamental changes in the law (FairTest Website).

The town has small class sizes, high graduation and college attendance rates, and strong test scores. Many residents had been shocked to learn that six of Brookline’s nine schools had failed to make AYP for the previous year.

Effective change will require more vigorous public education coupled with grassroots resistance that is in turn linked to the growing network of organizations seeking fundamental changes in the law.

To get started, here are some concrete actions that parents, educators and other concerned citizens can take:

1. *Hold a forum with parents and teachers about the law - and invite policymakers to listen, not to lecture you.* Invite a range of community groups to help plan, publicize, fund and facilitate the event: PTOs, parent groups for special education, gifted and bilingual education groups, teachers unions, local education foundations, civil rights organizations, the chamber of commerce, realtors, anyone who has a stake in the health of the local public schools. Invite the press to give advance publicity and cover the event. Videotape the event so it can be played at future meetings or on local cable television (or ask them to tape and broadcast it). Collect names and contact information from attendees who may want to stay involved in the issue and create or join a local education reform organization. Allow time for questions as well as a chance for the audience to express their observations and concerns about the law. Be sure to have sufficient written material for people to take home and to pass on to family, friends and colleagues.
2. *Circulate information about the law and its consequences.* Obtain (and edit if needed) or write materials to be used as widely as possible to educate people about the law and its flaws. See the FairTest website, <http://www.fairtest.org>, for sample material.
3. *Persuade your local school board and the state association of school boards to pass a resolution critical of NCLB; get the resolution endorsed by a variety of other groups.* There is a growing list of such resolutions that have been passed around the country (NEA, 2004) that can be used as models. Or use a petition to guide the crafting of a resolution (see below). Use local networks of public school supporters and activists to organize groups of community members to speak in favor of such resolutions during public comment periods of local government agencies. If your local board has taken such action on its own initiative, make sure they know you appreciate their action and would like to help them take it further.
4. *Circulate an NCLB-reform petition.* Petitions serve several valuable functions: They can be used to educate the public about the issue, they provide a vehicle for organizing parents and others to

Hold a forum with parents and teachers about the law - and invite policymakers to listen, not to lecture you.

Persuade your local school board and the state association of school boards to pass a resolution critical of NCLB; get the resolution endorsed by a variety of other groups.

continue pushing for change, and of course they demonstrate to policymakers that momentum is building to overhaul NCLB. If you circulate a petition, if possible, have written material and contact information that petitioners can hand out. FairTest is working with other groups to develop a petition to be circulated nationally. Check the FairTest website for news about the petition (www.fairtest.org/nattest/bushtest.html).

5. *Use NCLB's requirement for parent involvement to challenge the worst aspects of the law.* Under the law, districts must create parent advisory councils to provide for parent involvement in school improvement plans and other aspects of the law. These councils offer the chance for parents to speak out on NCLB's negative consequences and call for change.
6. *Develop capacity and prepare for an extended struggle.* Public forums and petitions can be valuable for educating people. It is vitally important, however, to use multiple vehicles. Parents and teachers need a variety of opportunities to exchange information at formal and informal occasions about the effects of NCLB and other test-based accountability programs. Advocates need to write letters and opinion columns for community newspapers. It is helpful to develop a network of people who can speak effectively with local reporters.

More confrontational actions can also be effective. In some states, parent and student refusal to participate in testing has garnered public and media attention. These boycotts also become an occasion for public education and discussion. The high stakes attached to the tests can make refusal a difficult choice, but a very powerful one, particularly if a coalition of parents and educators publicly support such actions. Rallies, demonstrations, and public picketing have also been used effectively. In a few cases, non-binding ballot resolutions also have been used.

Educating and organizing people around NCLB will create and expand a lasting network or organization devoted to authentic education reform. By marshalling the growing awareness and opposition to NCLB's punitive and destructive approach, there is an opportunity to develop the power to persuade political leaders that now is the time to rewrite the law. Only an informed and energized public can take the many forms of action needed to win fundamental changes.

Use NCLB's requirement for parent involvement to challenge the worst aspects of the law.

The high stakes attached to the tests can make refusal a difficult choice, but a very powerful one, particularly if a coalition of parents and educators publicly support such actions.

References

- Albuquerque Tribune* [NM]. March 3, 2004. "Let's Exempt N.M.'s Schools from Bad Law." Available online at http://www.abqtrib.com/archives/opinions04/031604_opinions_edtue.shtml.
- Brookline School Committee. January 15, 2004. Brookline School Committee Position on "No Child Left Behind." Available online at http://bec.brookline.mec.edu/Brooklineschool_com/Resolutions.html.
- Bruce, A. March 14, 2004. "Education Law Sparks Mounting Criticism," *Charleston.net*.
- Council of Chief State School Officers. January 28, 2004. "NCLB Implementation Issues and Opportunities for Action." Available online at www.ccsso.org/content/pdfs/4-2NCLBImplementationIssuesandOpportunities.doc.
- Dillon, S. February 22, 2004. "Bush Education Officials Find New Law a Tough Sell," *The New York Times*.
- FairTest *Examiner*. Winter-Spring 2003. "MA School Boards Defy State." Available online at http://www.fairtest.org/examarts/Winter-Spring%2003%20double/MA_School_Boards.html
- FairTest *Examiner*. Winter 2000-2001. "Massachusetts Exam Under Attack." Available online at <http://www.fairtest.org/examarts/Winter%2000-01/Mass%20Exam%20Under%20Attack%20w01%20p4.html>
- FairTest Website. Fact sheets on NCLB are on the national page <http://www.fairtest.org/nattest/bushtest.html> (Position Statement, "NCLB After Two Years"); other fact sheets on testing are at <http://www.fairtest.org/facts/main.htm>. For material on working with the media see <http://www.fairtest.org/arn/ARNmediaGuide.html>. For information on state activities, see the Assessment Reform Network pages <http://www.fairtest.org/arn/arn.htm>.
- Honolulu Advertiser* [HI]. January 6, 2004. "School 'Reform' Cannot Be Left to Schools Alone" (Cited in NEA, 2004).
- Lansdale Reporter* [PA]. November 11, 2003. "Education Cannot Be One Size Fits All" (Cited in NEA, 2004).
- Massachusetts Association of School Committees. November 12, 2003. "Delegate Assembly Adopts Eight Resolutions," MASC Website. Available online at http://www.masc.org/con02_nomframe.asp.

- Metropolitan Nashville Public Schools. September 2003. "Board Suggests Amendments to 'No Child Left Behind Act.'" Available online at http://www.mnps.org/Director/Board_folder/Board_NCLB_actions.html
- National Conference of State Legislatures (NCSL). March 30, 2004. "Task Force Formed to Examine Federal No Child Left Behind Act." Available online at <http://www.ncsl.org/programs/press/2004/pr040330.htm>
- NCSL. August 26, 2003. Letter to Mr. David Dunn, Special Assistant to the President for Domestic Policy.
- National Education Association (NEA). February, March 2004. "Ten Changes Needed to Improve 'No Child Left Behind.'" Available online at <http://www.nea.org/lac/esea/tenchanges.html>.
- NEA, April 16, 2004, "The Growing Chorus of Voices Opposing Provisions of the No Child Left Behind Act." Available online at <http://www.nea.org/esea/chorus1.html>.
- National School Boards Association. February 2004. "Ensuring the Goals of No Child Left Behind." Available online at <http://www.nsba.org/site/docs/33100/33080.pdf>.
- Neill, M. 1997. *Testing Our Children: A Report Card on State Assessment Systems*. Cambridge, MA: FairTest.
- Palm Beach Post* [FL]. January 3, 2004. "Policy Favors Schools That Leave Out Diversity" (cited in NEA, 2004).
- Public Education Network. 2004. "2004 Survey of Public Opinion: Learn. Vote. Act. The Public's Responsibility for Public Education." Available online at http://www.publiceducation.org/portals/Learn_Vote_Act/default.asp.
- Results for America. 2004. "Poll of Parents on 'No Child Left Behind.'" Available online at <http://www.resultsforamerica.org/education/#>
- School Board News*. March 16, 2004. "States Urge Congress to Fix — or Repeal — NCLB."
- School Board News*. April 13, 2004. "Public Supports Schools."

Appendix 1: What's in a Name?

At first, it seemed the best thinking that went into the 2001 reauthorization of the Elementary and Secondary Education Act was in devising a title that virtually everyone could endorse. Bush's team took a slogan from the Children's Defense Fund, "Leave No Child Behind," and changed it to the passive voice: "No Child Left Behind." The title implies a determination to rectify our country's long history of inequities to ensure that every child, regardless of economic status, race or disability, would be brought forward by this historic effort at public school reform.

However, as more and more constituencies have concluded that the law is underfunded, unworkable, or both, parodies of the name have proliferated. One measure of the depth and breadth of opposition to NCLB may be in the rapid accumulation of alternate, mocking versions of the title.

Perhaps the most popular and accurate one is "Leave No Child Untested," spoken by Wisconsin Superintendent of Instruction Elizabeth Burmaster in 2001 and repeated many times thereafter by numerous others (Borsuk, 2002)

Her comment appeared to rattle U.S. Education Secretary Rod Paige. Paige replied, "We don't want any child to be left untested. That is a fact." But he said testing was a key to holding schools and teachers accountable and to improving the results for every child, and he said the law contained much more than the testing requirement.

A succinct version is "No Child Left," the title of a web site that describes itself as "a site devoted to a sound approach to school improvement." The site's publisher, Jamie McKenzie, Ed.D., is a former teacher, principal and superintendent deeply troubled by the way test-based "reforms" like NCLB tend to narrow and impoverish children's schooling. The title of the site reflects his strong belief that high-stakes testing and an emphasis on punishment, negative labels and threats can do great damage to schools and children (McKenzie, 2004).

Despite the law's old-fashioned test-and-punish approach, it would be wrong to see this merely as a debate between liberals and conservatives. Opposition to the law (and sendups of the title) come from all

One measure of the depth and breadth of opposition to NCLB may be in the rapid accumulation of alternate, mocking versions of the title.

Despite the law's old-fashioned test-and-punish approach, it would be wrong to see this merely as a debate between liberals and conservatives.

over the political spectrum. An article from the right-wing Eagle Forum in January 2002 was entitled, “Will New Education Law Leave Every Child Behind?” (*Education Reporter*, 2002).

The article complained that instead of holding school districts accountable to parents, the law would hold them accountable only to the federal government, a point that has not necessarily been acknowledged by NCLB’s liberal supporters.

Former presidential contender Howard Dean, while critiquing the law from a different perspective, used similar language to the Eagle Forum. “I actually call this act the ‘leave every child behind’ act” (CNN, 2002). Dean’s stump speeches often focused on the failings of NCLB, so he developed a collection of alternate monikers for the law: “School boards call it ‘No School Board Left Standing.’ ... Teachers call it ‘No Behind Left’” (Robelen, 2003).

AFL-CIO President John Sweeney said education would be among organized labor’s top issues in 2004. He charged that the president’s proposals to cut taxes show he cares more about taking care of business executives than educating children. “While he is saying No Child Left Behind,” Sweeney said, “it is more like ‘No CEO Left Behind’ with his tax policies” (Dodge, 2003).

Eric Crane, a school committee member in Danvers, Massachusetts, was livid when he learned that remediation programs for the state test had been cut while test mandates were being increased. “I’m starting to think that NCLB stands for ‘No Cash Left Buddy,’” he said (Puffer, 2003).

Joyce McGreevey, writing in Salon, called it, “All Testing All the Time.” She said that proponents “believe that keeping teachers too busy to actually teach will better enable them to prove that they are, as the law now demands, ‘highly qualified teachers.’”

Familiarity with NCLB continues to breed contempt, or at least satire. Here are just a few of the many variations: No School Left Unpunished, No Teacher Left Standing, No Child’s Behind Left, No School Left Unlabeled, No Student Left Unrattled, No Parent Left Unsettled, No Test-Prep Company Left Behind, No Child Left Better, No Child Left with a Mind.

“While he is saying No Child Left Behind,” Sweeney said, “it is more like ‘No CEO Left Behind’ with his tax policies.”

Proponents “believe that keeping teachers too busy to actually teach will better enable them to prove that they are, as the law now demands, ‘highly qualified teachers.’”

Perhaps the task before those who seek sound school improvement, as Dr. McKenzie puts it, is to marshal the passion and energy fueling all this political expression and come up with an education reform law worthy of its name.

Borsuk, A. June 21, 2003. *Journal Sentinel*, EducationNews.org.

Bracey, J. October 2003. "The 13th Bracey Report on the Condition of Public Education," *Phi Delta Kappan*.

CNN's "Inside Politics." June 3, 2002. Transcript available online at <http://www.cnn.com/TRANSCRIPTS/0206/03/ip.00.html>.

Dodge, R. August 11, 2003. "Educators Fear Loss of U.S. Dollars Will Hinder No Child Left Behind," *Dallas Morning News*.

Eagle Forum. January 2002. "Will New Education Law Leave Every Child Behind?" *Education Reporter*. Available online at <http://www.eagleforum.org/educate/2002/jan02/hr1.shtml>.

McGreevy, J. August 11, 2003. "Less Teaching, More Testing," *Salon*. Available online at <http://salon.com/tech/col/mcgreevy/2003/08/11/education/print.html>.

Puffer, M. October 22, 2003. "State Cuts MCAS Grants," *Danvers Herald*.

Redovich, D. June 2002. "Just Another Big Con: The No Child Left Behind Act," *Journal Sentinel*.

Robelen, E. November 12, 2003. "On Trail, It's Dean vs. No Child Left Behind Act," *Education Week*.

Robinson, D. October 28, 2003. "No Teacher Will Be Left Standing," *Deseret Morning News*.

Truman, C. November 19, 2003. "What the School Scores Really Tell," *Lexington (KY) Herald-Leader*.

Perhaps the task before those who seek sound school improvement, as Dr. McKenzie puts it, is to marshal the passion and energy fueling all this political expression and come up with an education reform law worthy of its name.

VI. Authentic Accountability

“Educators today are besieged by a movement that demands higher and higher scores on standardized tests. Anyone who has looked carefully at these tests knows that they are loaded with trivia—questions that most successful adults cannot answer and would indeed scorn to answer. Our children are being fed intellectual junk food, and we would do well to insist on a healthier educational diet.”

- Nel Noddings, “War, Critical Thinking, and Self-Understanding,” *Phi Delta Kappan*, March 2004.

Based on the first two years of its implementation, the “No Child Left Behind” (NCLB) law must be overhauled. It should be replaced with new federal legislation that focuses on educational improvement based on rich goals and comprehensive student assessment, not on punishments in response to narrow standardized tests.

The escalating debate over NCLB appears to offer a choice between the “bad old days” of substandard schooling for too many students, or the “bad new days” of one-size-fits-all education aimed at meeting unrealistic test score targets. In truth, there is no need to choose between these poor options because better alternatives exist.

NCLB endorsed the laudable proposition that all children should have the chance to attain meaningful educational goals. The fundamental flaws in the law – its disastrous underfunding, reduction of education to preparation for narrow standardized tests, one-size-fits-all rigidity, punitive approach, and empowerment of distant bureaucracies – should not become excuses to abandon the goal of quality education for all children. The solution is not to punish, but to promote the use of methods that will improve education so as to really leave no child behind.

The current law includes some worthy elements. It allows for local assessments, including classroom-based assessments, as part of a state assessment system. It also requires multiple measures, use of improvement plans and professional development, disaggregation of data, and highly qualified teachers. Unfortunately, the positive elements are often swamped by the law’s fundamentally punitive structure. Some good elements, such as the call for measures beyond standardized test

The escalating debate over NCLB appears to offer a choice between the “bad old days” of substandard schooling for too many students, or the “bad new days” of one-size-fits-all education aimed at meeting unrealistic test score targets. In truth, there is no need to choose between these poor options because better alternatives exist.

Unfortunately, NCLB’s positive elements are often swamped by the law’s fundamentally punitive structure.

scores, are essentially ignored, and many of the particulars of the improvement process are poorly thought through.

Education, civil rights, parent groups and community organizations have been working with researchers to develop alternative models of accountability to guide revising the federal law. What follows owes much to those discussions, as well as to important work done in some states (particularly Nebraska and Maine), districts and many schools. Alternative models need not be invented from scratch. High-quality assessments already exist, as does research showing how their use can powerfully assist teaching and learning. Many educators and researchers have studied the problem of accountability and proposed reforms that should also inform any effort to craft revised legislation.

This chapter presents several approaches to thinking about accountability that focus on improving teaching and learning.

- First, a set of principles that should guide development and implementation of an improvement and accountability system.
- Second, a summary of the assessment and accountability model proposed for Massachusetts by the Coalition for Authentic Reform in Education (CARE), one that incorporates elements that should be included in a revised NCLB.
- Third, a discussion of Nebraska's assessment and accountability program, a concrete example of one state trying to make authentic accountability work and having some success.

The chapter concludes with a brief discussion of the politics of replacing the current NCLB with a very different model.

A. Principles for Authentic Accountability

FairTest is collaborating with education, civil rights, parent and community organizations, and researchers to develop a set of principles that can guide accountability programs. The portions of the draft principles included below are intended to spur further discussion; they are not a final product nor have they been endorsed by any other organizations. They do extend beyond testing and assessment to address fundamental issues of public schooling.

Alternative models need not be invented from scratch. High-quality assessments already exist, as does research showing how their use can powerfully assist teaching and learning.

FairTest is collaborating with education, civil rights, parent and community organizations, and researchers to develop a set of principles that can guide accountability programs.

Accountable to What Ends?

The key purposes of accountability are to inform the public – to give an accounting – of the status of the school or system; to provide information that can be used to improve teaching and learning; to ensure equity within the system; to strengthen democratic participation in governing schools; and to ensure that participants in the system carry out their responsibilities well.

Clearly there is much about schools, particularly those serving low-income and minority-group children, that must be improved. Therefore, the chief goals of accountability should be improvement and equity. Because quality public schools are widely understood as foundational to democracy, accountability procedures should strengthen, not undermine, participatory democracy. All the principles should be used to help guide participants in the system to do their work responsibly and well.

We propose four broad principles for the purposes of authentic accountability:

1. *Improvement.* Schools and districts must be accountable for using a range of measures of school health to make decisions and implement procedures over time that will improve the quality of schools and learning. Good teaching is fundamental. Since a primary purpose of accountability is to make schools better, professional development – particularly time for teachers to collaborate – must be a regular part of teachers’ paid work and must be aimed at improving practice.
2. *Equity.* Education systems can and should contribute to closing the race and class achievement gaps and to overcoming the consequences of poverty and racism. Gaps must be closed on the significant academic, personal and social outcomes that society wants for its children – not only on standardized tests – and on the social and school “inputs” that powerfully shape school achievement. Schools may need to create links with social service and health agencies to provide, for example, vision or dental care. In addition, schools should gather and publicize information about unmet social needs that hinder student achievement. Children who need more should be provided with more: equity does not mean the same for all, it means that all children receive what they need to develop fully.

The chief goals of accountability should be improvement and equity.

Education systems can and should contribute to closing the race and class achievement gaps and to overcoming the consequences of poverty and racism.

3. *Democracy.* Control over and responsibility for schools must be grounded in sound principles of participatory democracy. Accountability systems, therefore, must promote the informed involvement of key actors in the education system: parents, students, educators, and members of the local community first of all. To further strengthen democracy as well as promote equity and overall achievement, government and education systems should be accountable for promoting, expanding and strengthening schooling that is integrated by race and class.
4. *Informing the public.* The public deserves accurate information about the functioning, successes and problems of public education, focusing on the various aspects of schooling that are of major concern. For example, in addition to information on achievement (which must include more than test score data), the public needs to know if schools lack basics like well-equipped and staffed libraries, art supplies and science labs, and clean bathrooms.

Accountable for What?

Accountability must be based on a shared vision and goals for education and schools, that is to say, on agreement about what schools should be and do. The larger community must participate in setting the basic goals and purposes of the educational system and evaluating how well those goals have been met. Because a shared vision may not be present, processes must be established to enable communities to come to agreement or to allow differences to co-exist. To meet this purpose, we propose the following five principles:

1. *Priorities.* The shared vision should establish priorities for the following: academic and other formal learning; students' physical and emotional well-being; schools' social environment; and how well schools prepare students to participate in our democracy, be lifelong learners, and make a good living. Assessment information used in accountability must focus on those areas deemed most important, not only on those areas that are easiest to measure with inexpensive tools, such as standardized tests, though such tools have a place in the accountability process.
2. *Resources.* Government must be held accountable for providing education systems, including schools and pre-schools, with adequate resources to meet agreed-upon priorities. This includes

Equity does not mean the same for all, it means that all children receive what they need to develop fully.

In addition to information on achievement (which must include more than test score data), the public needs to know if schools lack basics like well-equipped and staffed libraries, art supplies and science labs, and clean bathrooms.

the money to hire good teachers and ensure continuing professional development. It also includes money to provide small classes, books, and technology and supplies in a comfortable, clean and hospitable environment in order to ensure that all children receive an adequate and equitable opportunity to learn. Resources for other policies and programs known to contribute to important outcomes, such as pre-school or health care, must also be provided. Schools and districts should be accountable for using their resources fairly and effectively.

3. *Student learning.* Education systems should be accountable for ensuring that all students learn those essential things society agrees all should learn (i.e., academic standards) and for enabling all students to pursue areas of individual interest and talent. Assessments of academic, vocational or other formal learning must promote, measure and provide useful feedback on conceptual understanding and the ability to use knowledge and create rather than primarily procedural, factual or surface learning. They must include all important content areas of learning and be congruent with current knowledge about how students learn. Graduation rates and post-secondary success should be included in accountability reports, broken out by key demographic groups.
4. *Student well-being.* Students are happier and achieve more in environments that are hospitable and welcoming and where students feel empowered, challenged, motivated and supported. To hold schools accountable for establishing supportive and caring learning environments for all children and for ensuring students' physical and emotional well-being, there must be evidence that illuminates these aspects of schools' environments.
5. *Inclusion.* The progress and well-being of all students must be accounted for. All accountability data should be broken out by major demographic categories. Inclusion also implies respect for the diverse experiences and cultural backgrounds of students and communities.

Accountable to Whom?

Accountability must be mutual and reciprocal. An accountability system must define appropriate expectations for participants in the system (e.g., schools, districts, the state and federal governments, as well as students and teachers).

Government must be held accountable for providing education systems, including schools and pre-schools, with adequate resources to meet agreed-upon priorities.

Graduation rates and post-secondary success should be included in accountability reports, broken out by key demographic groups.

1. *Higher levels of government* are responsible for providing sufficient resources to ensure adequate and equitable opportunity to learn; for safeguarding civil and human rights; for monitoring local systems; for analyzing research and practice to determine what works best in what circumstances; for disseminating knowledge; for providing additional support as needed; and for intervening in localities when necessary.

States can define core areas for learning, though specific standards as well as curriculum and instruction can be left to districts and schools. State agencies should intervene when localities are unable to provide a high-quality education even when they have reasonable resources. Governments are accountable for conducting business with transparency and substantial educator, parent and community input.

2. *Local schools and districts and their communities* must be the primary authorities in the accountability process. Schools are first of all accountable to their students, their parents and the local community. Local accountability involves active participation and shared power among key actors. Schools and districts also are responsible to the general public and the state.
3. *The accountability structure* delineates roles and responsibilities of the state, districts, and schools, gathers evidence as to how well they have been carried out, and includes means to ensure change and improvement where necessary.

Accountable by what means?

The means used to implement accountability can support or undermine underlying goals and overall school quality. The trend in education policy, exemplified especially by NCLB, has been to combine narrow measures with high stakes, thereby damaging schools' capacity to meet larger goals and often undermining the quality of education.

To ensure that accountability methods support comprehensive accountability goals:

1. *Use multiple forms of evidence.* Accountability requires the use of multiple forms of qualitative and quantitative evidence from both academic and non-academic areas to assess students, schools and districts and determine how to make improvements. All students must be assessed and evaluated with a range of

State agencies should intervene when localities are unable to provide a high-quality education even when they have reasonable resources.

Schools are first of all accountable to their students, their parents and the local community.

appropriate tools and methods. No important academic decision about a student, a teacher, an administrator, a school or a district should be made solely on one type of evidence, such as standardized test scores. Multiple forms of evidence may include teacher evaluations of student achievement, portfolios as evidence of student work, final projects presented to a panel of community members, etc. Scores from several standardized tests do not constitute multiple forms of evidence.

2. *Assess a set of key factors that are known to predict school and system success.* These “predictive” or “formative” indicators include in-school factors such as strong classroom assessment and professional development for teachers, and out-of-school factors such as health care, housing, nutrition, and availability of high-quality pre-school; whether educators are using information in a reasonable way to improve teaching, learning and school quality; and whether the state and federal governments are providing positive support in these areas. Practices proven to inhibit high achievement or harm school quality, such as tracking, retention and lower expectations for some groups of students, should be identified and analyzed.
3. *Use skillful feedback to improve student outcomes.* Research has demonstrated that skilled use of feedback to students (“formative assessment”) is among the most powerful tools teachers have to help students learn. For assessment to be most helpful and guide further instruction, it must be comprehensive and regular enough to provide fine-grained information about each student, and the student must understand and apply the information. Most assessment, therefore, must be classroom-based and used by well-prepared teachers. Schools and districts must ensure that all teachers become skilled at this kind of assessment. Standardized exams should supplement, not supplant or overpower, classroom assessment.
- 4) *Use interventions sparingly and carefully.* Interventions from higher levels of government must focus on providing useful assistance and include harsher measures only as a last resort. Intervention should focus on factors that can produce significant improvement, including effective professional development, active parent involvement, high-quality classroom assessment, and smaller class sizes.

Accountability requires the use of multiple forms of qualitative and quantitative evidence from both academic and non-academic areas to assess students, schools and districts and determine how to make improvements.

Practices proven to inhibit high achievement or harm school quality, such as tracking, retention and lower expectations for some groups of students, should be identified and analyzed.

If a school or district has taken steps that plausibly will lead to desired improvement, it must be allowed time for those changes to take effect. During that time, improvement efforts must be monitored using a range of evidence to determine if implementation of reasonable changes is proceeding well and schools are able to use information to effectively adjust their improvement efforts. If a school or district is unable to improve despite assistance, then a higher level of government should intervene.

The little research that exists suggests that there is no significant evidence that sanctions such as removing the principal and key staff, privatizing school or system management, making the school a charter school, or having the state run the school or district are effective or create improvement. Such sanctions therefore should be taken as a last resort, with sufficient support and resources to increase the likelihood of success, and with careful monitoring of progress. Such strong interventions should be consistent with these principles.

B. Massachusetts CARE: Call for an Authentic Statewide Assessment System

The Coalition for Authentic Reform in Education (CARE) has proposed an alternative, authentic assessment and accountability plan for Massachusetts (the full plan is available online at <http://www.fairtest.org/care/accountability.html>). It contains many features that should be included in a revised version of NCLB. It would replace the current state reliance on one set of tests, the MCAS exams. In collaboration with the Massachusetts Teachers Association, CARE has submitted its proposal as legislation. A modified version has been developed by parent, community, civil rights and education reform organizations in Chicago, and it is called the New ERA plan.

The CARE plan is based on five key points:

1. To know how students are learning, look at the work they do and at their teachers' assignments.
2. For effective staff development and school improvement, it is essential that teachers review student work together and use that information to think about improving teaching.
3. Local schools know their students best.
4. Local communities must play the primary role in evaluating local schools.

Standardized exams should supplement, not supplant or overpower, classroom assessment.

To know how students are learning, look at the work they do and at their teachers' assignments.

5. The state's job is not to make decisions about individuals but to ensure that schools are educating all children well and to provide the necessary resources to enable schools to do so.

The CARE plan builds on Massachusetts' *Common Core of Learning*, a brief statement of essential learning goals for all children. CARE calls for expanding the Core to define "core competencies" that are leaner than the state's long, detailed, and complicated curriculum frameworks, leaving these core competencies to be filled out by districts and schools in adopting standards and then in shaping curriculum and instruction.

The key elements of the CARE proposal are as follows:

1. **Local authentic assessments.** These will be based on the new "competencies" and a school's own goals. Each school and district will have an assessment and accountability plan — approved by the local school council, the state and the district — which explains how it will assess students, how decisions such as graduation and grade promotion will be made, how it will use information about student work to improve teaching, and how accountability information will be reported to parents, students, teachers, the community and the state. Graduation will be decided by the school, not by the state.
2. **Limited standardized testing** in literacy and numeracy only. These tests will not be used to make decisions about students but will be one source of data about individual, classroom and school performance.
3. **School quality reviews.** Every four to five years, each school will do a detailed self-study. Then an independent, expert team will conduct a several-day visit to the school, interviewing students, educators, and parents, sitting in on classes, looking at examples of student work, etc. The team will present a detailed report to help guide the school in making further progress. The teams might be organized by the Department of Education or developed by the regional accreditation association.
4. **Annual school reporting.** Each school will report on progress or lack thereof toward its goals and the state's *Common Core of Learning*, and how it is using information about teacher assignments and student work to improve the school. The report will be based on the local assessments and include standardized test results. Outcomes by race and ethnicity, gender, low-income

The state's job is not to make decisions about individuals but to ensure that schools are educating all children well and to provide the necessary resources to enable schools to do so.

Every four to five years, each school will do a detailed self-study. Then an independent, expert team will conduct a several-day visit to the school, interviewing students, educators, and parents, sitting in on classes, looking at examples of student work, etc.

status, special needs, and limited English proficiency will be included. So will other information about the school, such as attendance, promotion and dropout data; survey results (such as school climate surveys); teacher qualifications; and resource availability. Data will be reviewed by the local school council, parents and other community members, the district, and the state. When needed, the state or district can send in teams to verify the accuracy of a school's report.

In this accountability system, much more information will be available than is provided by state testing programs. No one test will determine the fate of a student or a school. The plan builds in a process of continuous improvement. The state will have sufficient information to intervene in a school or district that has adequate resources but does not perform well and does not improve.

Public participation is essential to the CARE and New ERA plans. A school's community comes together to evaluate the information and improvement plans. Teachers, administrators, parents and students can openly discuss the successes and problems and come to agreements on where improvements are needed. In some Chicago schools, the local school councils, which have a parent majority already engage in such discussions.

The CARE and New ERA plans call for intervention in schools that clearly demonstrate they are not succeeding, according to multiple measures, for significant portions of their students. The first step is an investigation, particularly through the quality review process. A review must include relevant factors that may be beyond a school's control, such as family poverty or student mobility, the resources a school has, what it does with its resources, and how it might use resources better.

If needed, the district or state should provide carefully targeted assistance. If a school still does not make progress even with assistance, stronger interventions should take place. However, too little is known about how to make such interventions succeed. Therefore, states should develop and implement the stronger actions with caution and keep very close track of what does and does not work.

Public participation is essential to the CARE and New ERA plans. A school's community comes together to evaluate the information and improvement plans.

The CARE and New ERA plans call for intervention in schools that clearly demonstrate they are not succeeding, according to multiple measures, for significant portions of their students.

C. Nebraska

“The approach to standards, assessment and accountability in Nebraska is unlike that in any other state. Nebraska’s STARS (School-based Teacher-led Assessment and Reporting System) is not based on external mandates and compliance but relies instead on the professional judgment of teachers about whether their students are learning. Commissioner Doug Christensen has remained steadfast in his belief that ‘decisions about student learning reside in the classroom where learning occurs, not in the legislature, the governor’s office, or the department of education.’”

- Pat Roschewski, “Nebraska STARS Line Up,” *Phi Delta Kappan*, March 2003.

Nebraska Commissioner of Education Doug Christensen emphasizes that successful accountability must be done with, not to, teachers. Under his leadership, Nebraska has developed a unique state approach that provides valuable lessons for using assessment and accountability constructively.

While Nebraska has a set of standards, it allows districts to develop their own standards, provided that the state determines the district standards are high quality. Currently, a combination of norm- and criterion-referenced assessments are required for evaluating students in grades 4, 8, and 11 in mathematics, reading/writing, science, and social studies. The norm-referenced tests (NRTs) must be selected from a state list. In addition, all students in grades 4, 8, and 11 participate in a statewide writing assessment. The NRTs and writing test, however, are not part of the state accountability system, though the results are collected and reported. Accountability is based on the districts’ criterion-referenced assessments.

Four years into operation, the Student-based, Teacher-led Assessment and Reporting System (STARS) requires school districts to develop local assessment plans that are aligned with state (or district) learning standards. The STARS plan uses multiple assessment measures, as in the CARE and New ERA plans, rather than relying on a single test. Districts can develop their own instruments (which may include classroom assessments such as observations, portfolios, and rubrics) or they may purchase them from commercial publishers. These assessments and results are used for the state assessment and accountability system. Nebraska thus has a system of local assessments.

Nebraska Commissioner of Education Doug Christensen emphasizes that successful accountability must be done with, not to, teachers.

Nebraska has developed a unique state approach that provides valuable lessons for using assessment and accountability constructively.

One reporter (Dell'Angela, 2004) observed assessment in action in one suburban Nebraska school:

At Portal Elementary in La Vista, second grader Macy Morrison can see for herself that she's making progress. She has been taking tests since school started. By the end of the year, her teacher will send the district 33 measures of Macy's progress in reading, writing and math.

During a recent visit, Macy was reading an "Arthur" book into a microphone on a computer. This test measures Macy's fluency—a rare example in which speech is actually measured for state standards—and when she's finished she knows exactly what she should do to improve.

"My expression was just right, but I'm still getting there on my smoothness because I had a lot of stops," Macy said, clicking to a bar chart of her progress during second grade.

Districts must follow six criteria in designing their assessment plans: assessments reflect state or local standards; students have an opportunity to learn the content; assessments are free from bias; the level is developmentally appropriate for students; there is consistency in scoring; and mastery levels are appropriate. The state recently issued new regulations requiring districts to ensure that their assessments meet the breadth and depth of the standards.

Local assessment portfolios are submitted to the Nebraska Department of Education for review by an independent panel that rates the portfolios' quality. If a district's assessment obtains a rating of very good or excellent, it need not resubmit the portfolio for four years (e.g., a reading assessment approved in 2003 will be resubmitted in 2007). District Assessment Portfolios that are not approved are modified and resubmitted the following year. All assessment systems, however, are expected to be regularly improved, and the state is establishing a process for providing feedback to districts.

In essence, Nebraska has created standards for local assessments, a means to evaluate them in light of the standards, and a structure for ensuring that every district's assessments improve. If each district has strong standards and a high-quality assessment program, then it is reasonable to assume that if a teacher determines a student has reached a particular learning level, that determination is correct.

The STARS plan uses multiple assessment measures, as in the CARE and New ERA plans, rather than relying on a single test.

Local assessment portfolios are submitted to the Nebraska Department of Education for review by an independent panel that rates the portfolios' quality.

Chris Gallagher (2004), who is reviewing the STARS system, reported on its positive consequences:

Meanwhile . . . the kids in Palmer do Community Math, skillfully solving complex real-world problems they have solicited from their neighbors and parents. The kids in Cedar Bluffs do the “Platte Attack,” writing wonderful sandbar poetry and keeping science journals on the banks of the river that runs through their rural town. The kids here in Lincoln develop math portfolios, reflecting on and documenting their learning every day. The kids in Heartland build immigrant trunks and capably present their family heritage projects to a full auditorium every year.

An F [in “standards and accountability” from *Education Week’s* annual rankings] means that Nebraska continues to buck the high-stakes, test-’em-’til-they-drop mentality. It means that in Nebraska, assessment continues to be driven by instruction, rather than the other way around. It means that in Nebraska, as one teacher aptly puts it, having standards “does not make us all ‘standard.’”

The Buros Center for Testing at the University of Nebraska has assisted the state and has reviewed district assessment portfolios (Fairtest *Examiner*, 2002). Buros staff found the districts generally produced strong assessments and are willing to improve. Jim Impara of Buros reported that most districts decided not to simply develop criterion-referenced exams, but to use classroom-based assessments that could have a more positive impact on teaching and learning. This requires building district capacity to train teachers and ensure high-quality classroom instruction.

Each year, districts send to the state their students’ results on criterion-referenced assessments in the required grades, and if needed submit their assessment portfolio for (re)approval. The state is piloting an electronic portfolio for districts to use in submitting their materials. Districts are evaluated by separate panels on both elements (assessments and achievement), as can be seen on the state website. For districts in which classroom-based evidence is a key component of the assessment system, the classroom-based evidence is used by the state to determine how well districts are enabling students to meet the standards.

“An F [in ‘standards and accountability’ from Education Week’s annual rankings] means that Nebraska continues to buck the high-stakes, test-’em-’til-they-drop mentality.”

- Chris Gallagher

Classroom-based evidence is used by the state to determine how well districts are enabling students to meet the standards.

State leaders such as Doug Christensen and Pat Roschewski, Director of Statewide Assessment, conclude from extensive and continuing discussions around the state that the process has enabled teachers to learn a great deal and to create systems for talking with one another within and across districts. Those conversations have been the basis for improving the assessments. Reviews, such as those conducted by Buros and by Gallagher (2004) of the University of Nebraska, find that the assessment systems are having a positive effect on teaching and learning. Christensen reports district superintendents have come to support the program because children are learning more. In its studies, the state has identified six promising practices, the first being a shared vision and goals.

Writing in *Phi Delta Kappan*, Gallagher (2004) summarized ten lessons learned from the experience of Nebraska. They include:

1. Teachers should be regarded as leaders of reform, not impediments to it.
2. Accountability systems must focus on developing capacity, not controls.
3. Accountability systems must foster commitment, not compliance.
4. Accountability systems must promote integration of school improvement and accountability efforts.
5. Accountability systems must risk complexity rather than demand simplicity.
6. Accountability systems must include *all* students.
7. Accountability systems must also leave no teacher behind.
8. Accountability systems must engage *all* stakeholders.
9. Accountability systems must keep pedagogy – teaching and learning – at the center.
10. Accountability must promote high-impact, not high-stakes, assessment.

Nonetheless, there have been problems with implementation in Nebraska, including the following:

- There is a substantial burden on teachers, the expectations for teachers are higher, and they have had to validate their assessment practices and contribute to the state assessment program. Thus, there is some resentment and resistance.
- Some districts initially made the process burdensome. Most of those districts have been able to go back and improve their earlier efforts. Roschewski noted that the effectiveness of the

Christensen reports district superintendents have come to support the program because children are learning more.

Accountability systems must keep pedagogy – teaching and learning – at the center.

district leadership has a direct connection to the success of the assessment system and the belief in its importance. State staff members are working with districts and the leaders within districts to solve this problem and have those districts revamp their assessment systems.

- Implementation has been uneven. Gallagher's review found substantial gaps in quality. Roschewski thinks that about three-quarters of districts have coherent systems with promising practices. All but about 12 percent, or 89 districts, meet state requirements. Of those 89, some are doing acceptably well but have not known how to document it; some have useful elements in place but have not yet constructed a coherent system; and a few districts are still at square one, often because they assumed "this too shall pass" and did little. Gallagher found instances of "insufficient teacher participation in STARS across grades and curricula." In addition, he found too little engagement by parents and the community outside the school.
- Professional development takes time and resources. The state is expanding resources for practicing teachers, including summer institutes to review assessments and an 18-hour training that leads to a certificate in classroom assessment. The primary focus has been to create learning teams in schools, which, Gallagher reports, has led to a strong buy-in from teachers. By next year, all state universities will be required to include classroom assessment practices in their teacher education programs.

Nebraska leaders are enraged by NCLB: "I don't give a damn what No Child Left Behind says," Christensen said. "I think education is far too complex to be reduced to a single score. We decided we were going to take No Child Left Behind and integrate it into our plan, not the other way around. If it's bad for kids, we're not going to do it" (Dell'Angela, 2004).

Nonetheless, NCLB has had an impact on Nebraska. Roschewski reports that Nebraska has a dual assessment system, state and federal, as do many states. NCLB requires state assessments in reading and math in grades 3-8 and once in high school by the 2005-06 school year. Nebraska has informed districts that while they can use a norm-referenced test in grades 3, 5, 6, and 7, they also can expand their district assessments to incorporate those grades—which is what most are doing. Most districts have already identified local standards for those grade levels and have developed accompanying assessments for measuring

The primary focus has been to create learning teams in schools, which, Gallagher reports, has led to a strong buy-in from teachers.

"I don't give a damn what No Child Left Behind says, I think education is far too complex to be reduced to a single score."

- Doug Christensen

them. The districts will not have to present the assessments for those new grades for separate state review, but will cross-reference them to the portfolios submitted for grades 4, 8 and 11. Districts that expand use of classroom assessments to report student achievement will move closer to having the whole school involved in the assessment, which is also a Nebraska goal. If districts use norm-referenced tests to provide evidence of success, however, those tests could come to have undue educational influence. And the question of NCLB sanctions looms.

As part of the state accountability system, the state has given some districts one year to improve their assessments and three years to increase the percentage of students at proficient or higher (depending upon where the district is falling short). Those who don't improve sufficiently will have their accreditation put on probationary status. The state is offering intensive help to those districts. This process, however, is distinct from the AYP process mandated by NCLB. The state will address the sanctions attached to AYP this coming year, but state leaders clearly believe NCLB is not a good fit with Nebraska's STARS program.

In sum, Nebraska is the state closest in approach to the Principles for Authentic Accountability as described in this chapter. While it is a small state, nothing Nebraska has done is impossible for a large, more urban state to accomplish. Maine is the other state that is close to Nebraska's approach; it, too, is developing a local assessment system. Maine, Rhode Island and Wyoming have established policies in which students will have to demonstrate proficiency on local assessments in order to graduate. Rhode Island specifies that multiple forms of assessment must be used by districts. If the state test is used, it cannot count for more than ten percent of a student's score.

D. Conclusion: A Political Note

The key to this approach will be to persuade policymakers that cooperating with educators and communities to steadily improve schools across a wide array of indicators is preferable to threatening teachers and schools for failure to make progress on a few indicators measured with narrow tests. It is preferable because it will lead to substantial improvement – provided adequate resources – in ways the test-and-punish approach cannot and will not.

While it is a small state, nothing Nebraska has done is impossible for a large, more urban state to accomplish.

The key to this approach will be to persuade policymakers that cooperating with educators and communities to steadily improve schools across a wide array of indicators is preferable to threatening teachers and schools for failure to make progress on a few indicators measured with narrow tests.

This means jettisoning fixed AYP requirements and the illusions of certainty associated with numerical test scores. It does not mean ignoring low-income and minority-group children or having low expectations. It does mean believing that most schools will improve with guidance, assistance, professional development for teachers, stronger parent and community involvement, and adequate resources. It means recognizing that a climate of threats and sanctions is ultimately counterproductive. It means establishing a more flexible but still-definite willingness and capacity to intervene - states in districts, districts in schools.

Many will be skeptical that policymakers at the federal and state levels will switch from test-and-punish to assess-and-assist. Nevertheless, pursuing this goal is necessary for two reasons:

First, because the current approach is so disastrous that it must be fundamentally changed. It is important to establish as a goal a new concept and practice of “accountability.”

Second, by having a clear goal, it is easier to determine which partial steps and compromises help move toward the goal and which do not. For example, softening the rigidities of AYP will help reduce destructive forms of pressure, but by itself that will not help schools restructure themselves for making continuous progress on a rich array of learning outcomes or for using multiple measures to assess progress. Adding resources to help teachers become better at formative assessment will contribute to improved teaching, but these will be quite limited if AYP remains tied to state tests. Thus, both changes make sense, but not in isolation. They need to be part of a coherent package.

Since states are in revolt in large part over the lack of funding for NCLB, it is reasonable to ask about the costs of this new approach. Evidence suggests, for example, that additional costs to the state of Nebraska’s reform efforts have not been particularly large – but this is in part because the real cost is the work teachers do in their schools and districts. As many advocates of assessment reform pointed out in the 1990s, much of the cost of “assessment reform” reflect time for teacher professional development and time to reorganize schools and districts to better support high-quality learning by all students. If the goal really is to leave no child behind, the education system must spend the money needed for professional development and system change. This alternative accountability approach supports high quality changes and would

The current approach is so disastrous that it must be fundamentally changed.

Softening the rigidities of AYP will help reduce destructive forms of pressure, but by itself that will not help schools restructure themselves for making continuous progress on a rich array of learning outcomes or for using multiple measures to assess progress.

not add much to the cost. Continuing on the path demanded by NCLB and many state accountability programs will have a much higher cost: many children will continue to be left behind.

Continuing on the path demanded by NCLB and many state accountability programs will have a much higher cost: many children will continue to be left behind.

References

Buros Institute: www.unl.edu/buros/

Dell'Angela, T. April 5, 2004. "Nebraska Shuns State Tests; Schools Get Leeway to Judge Progress," *Chicago Tribune*.

FairTest *Examiner*, Spring 2002, "Nebraska and Maine Assessment Models."

Gallagher, C. January 2003. "Turning the Accountability Tables: Ten Progressive Lessons from One 'Backward' State." *Phi Delta Kappan*. Available online at <http://www.pdkintl.org/kappan/k0401gal.htm> The reports by his team, which formed the basis for the article, are on the Nebraska website.

Maine, see www.state.me.us/education/homepage.htm.

Nebraska: www.nde.state.ne.us/starsdocs.html

Roschewski, P. April 2001. "Nebraskans Reach for the STARS," *Phi Delta Kappan*. The material on Nebraska also relied on conversations with Doug Christensen, Nebraska Commissioner of Education; Pat Roschevsky, Nebraska Department of Education; and James Impara, Buros Center.