

FairTest

National Center for Fair & Open Testing

NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure?

By Lisa Guisbond with Monty Neill and Bob Schaeffer

January 2012

<http://fairtest.org/NCLB-lost-decade-report-home>

P.O. Box 300204, Jamaica Plain, MA 02108
fairtest@fairtest.org 617-477-9792 www.fairtest.org

NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure?

By Lisa Guisbond with Monty Neill and Bob Schaeffer,
January 2012

<http://fairest.org/NCLB-lost-decade-report-home>

Contents

Introduction.....	1
Part I. The Record: NCLB's Promises Unmet.....	2
Part II. Sour Wine with New Labels.....	10
Part III. Real Reform Is Possible but Would Mean Setting a New, Evidence-Based Direction.....	15
References.....	18

NCLB's Lost Decade for Educational Progress: What Can We Learn from this Policy Failure?

By Lisa Guisbond with Monty Neill and Bob Schaeffer
January 2012

Ten years have passed since President George W. Bush signed No Child Left Behind (NCLB), making it the educational law of the land. A review of a decade of evidence demonstrates that NCLB has failed badly both in terms of its own goals and more broadly. It has neither significantly increased academic performance nor significantly reduced achievement gaps, even as measured by standardized exams.

In fact, because of its misguided reliance on one-size-fits-all testing, labeling and sanctioning schools, it has undermined many education reform efforts. Many schools, particularly those serving low-income students, have become little more than test-preparation programs.

It is time to acknowledge this failure and adopt a more effective course for the federal role in education. Policymakers must abandon their faith-based embrace of test-and-punish strategies and, instead, pursue proven alternatives to guide and support the nation's neediest schools and students.

The data accumulated over ten years make three things clear:

1. NCLB has severely damaged educational quality and equity, with its narrowing and limiting effects falling most severely on the poor.
2. NCLB failed to significantly increase average academic performance and significantly narrow achievement gaps. And,
3. So-called "reforms," such as the Obama Administration's waivers and the Senate Education Committee's Elementary and Secondary Education Act (ESEA) reauthorization bill, fail to address many of NCLB's fundamental flaws and in some cases will intensify them. These proposals will extend a "lost decade for U.S. schools."

Despite a decade's worth of solid evidence documenting the failure of NCLB and similar high-stakes testing schemes, and despite mounting evidence from the U.S. and other nations about how to improve schools, policymakers cling to discredited models. This is particularly tragic for families who hoped their children's long wait for equal educational opportunity might be ending. It is also tragic for our public education system, whose reputation has been sullied by promises not kept and expensive intervention schemes that do more harm than good.

It is not too late to revisit the lessons of the past ten years and construct a federal law that provides support for equity and progress in all public schools. With that goal in mind, this report first provides an overview of the evidence on NCLB's track record. Second, it looks at recent efforts at NCLB "reform" and what past evidence says about *their* likely outcomes. Finally, it points to alternative strategies that could form the basis for a reauthorized federal law that would improve all schools, particularly those serving our most needy students.

Because of NCLB's misguided reliance on one-size-fits-all testing, labeling and sanctioning schools, it has undermined many education reform efforts.

Despite mounting evidence from the U.S. and other nations about how to improve schools, policymakers cling to discredited models.

Part I. The Record: NCLB's Promises Unmet

NCLB's ten-year report card offers little cause for celebration, whether you judge the law narrowly on its own terms or look more deeply at its impact.

- *NCLB's own narrow gauges of progress reveal major shortcomings: growth on the National Assessment of Educational Progress (NAEP) has stalled, achievement gaps are stagnant, and predictions of widespread school "failure" are coming true.*
- *The curriculum has narrowed, test preparation has displaced broader schooling, cheating is rampant, there is too little help for schools in need, and NCLB has contributed to the growth of a pernicious school-to-prison pipeline.*
- *A narrow focus on testing and punitive accountability has caused policymakers to ignore the real educational consequences of child poverty, which has grown significantly in recent years.*

Growth Stalled, Gaps Remain

Instead of helping to create circumstances in which schools can provide a rich, well-rounded curriculum and address the needs of individual students, the law has pressed schools to narrow curriculum, teach to the test, and resort to deceptive and unethical ways to boost test scores. It has done so by defining student learning and school quality in the narrow terms of standardized exam results.¹

NCLB's chief yardsticks for measuring results are state standardized tests in math and reading administered annually in grades 3 through 8 and once in high school. The law designated NAEP tests as an independent yardstick. School leaders and teachers correctly feared that failure to meet state test targets could result in sanctions for their schools. With so much riding on the results, many schools turned to preparing students for these tests, ignoring other aspects of education.

Not surprisingly, scores on state-administered tests have shown greater growth than NAEP, on which scores have tended to stagnate. However, as benchmarks moved higher, stretching toward the goal of 100% proficiency, more and more schools in almost every state have fallen short. This is due in large part to the law's requirement that every one of multiple groups—race/ethnicity, low-income, English language learner and disabled—make "Adequate Yearly Progress" (AYP). In the 2010-2011 academic year, 48% of the nation's 100,000 schools failed to reach AYP benchmarks.

Instead of helping to create circumstances in which schools can provide a rich, well-rounded curriculum and address the needs of individual students, the law has pressed schools to narrow curriculum, teach to the test, and resort to deceptive and unethical ways to boost test scores.

As benchmarks moved higher, stretching toward the goal of 100% proficiency, more and more schools in almost every state have fallen short.

¹ FairTest's report, *Failing Our Children* (Neill, Guisbond & Schaeffer, 2004), explains the myriad ways high-stakes testing damages the quality of education and undermines individual opportunities. In doing so, it explained why NCLB was going to leave many children behind.

What about the backup measure? NAEP, too, is a standardized test, primarily multiple-choice with some short-answer questions. It has been particularly criticized for its flawed definition of “proficiency.” Nevertheless, it is a technically sound standardized exam, generating consistent scale scores from year to year, allowing their use as an independent yardstick to track whether and when improvements have occurred.

The latest NAEP results (NCES, 2011a,b) confirm trends identified over the past decade (FairTest, 2009). Overall, growth on NAEP was more rapid before NCLB became law and flattened after it took effect. For example, 4th grade math scores jumped 11 points between 1996 and 2003, but increased only 6 points between 2003 and 2011. Reading scores have barely moved in the post-NCLB era. Fourth grade scores increased just 3 points to 221 between 2003 and 2011, remaining level since 2007. In 8th grade reading, there was a meager 2-point increase, from 263 to 265, in that same period. Since the start of NCLB, gains have stagnated or slowed for almost every demographic group in both subjects and both grades.

Overall, growth on NAEP was more rapid before NCLB became law and flattened after it took effect.

Figure 1. Trends in fourth- and eighth-grade NAEP math average scores



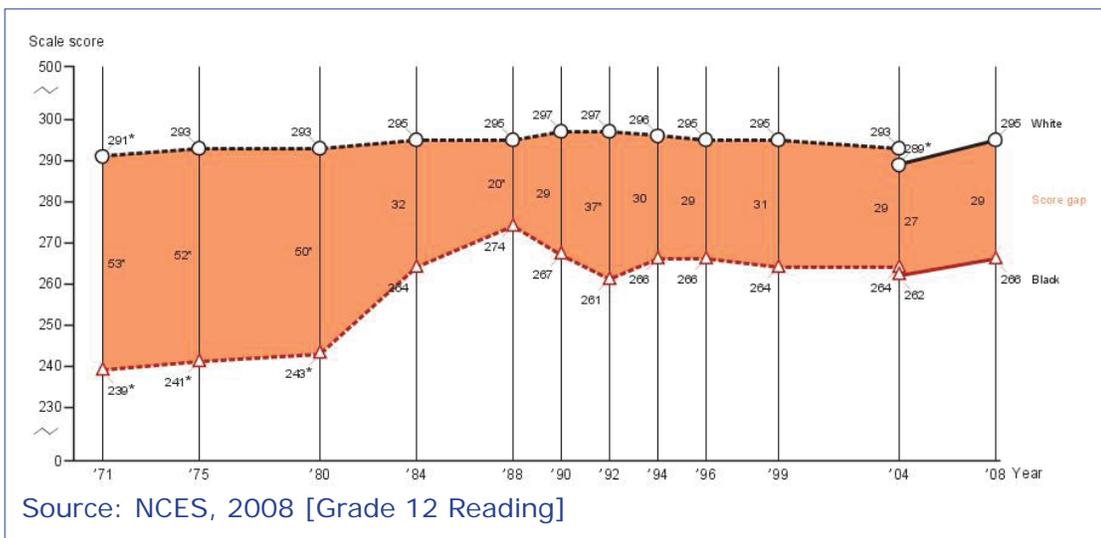
Source: NCES, 2011a

As a result, gaps between groups remain large, despite the hope that NCLBs exposure of these gaps would motivate successful efforts to close them. In fact, gaps have remained mostly stagnant for most groups of students at both grade levels in both subjects. For example, in 8th grade math, the large gap between Whites and Blacks remained at 32 points from 2007 to 2009, closing by just one point in 2011. In 8th grade reading, Wisconsin is the only state that narrowed the gap between Whites and Blacks between 1998 and 2011 and only two states, Alabama and California, narrowed the gap between Whites and Hispanics.

In 8th grade math, the large gap between Whites and Blacks remained at 32 points from 2007 to 2009, closing by just one point in 2011.

Columbia University Professor of Sociology and Education Aaron Pallas (2011) looked at changes in the performance of White, Black and Hispanic students in every state on 4th and 8th grade reading and math between 2003 and 2011. He concluded that NAEP “provides no evidence that states can meet the laudable goal of convergence of student-subgroup performance at a significantly higher level of academic proficiency than is currently observed. No state over the past eight years has succeeded in doing this in the way that NCLB demands” (Pallas, 2011).

In fact, long-term NAEP trends show just one period in which achievement gaps narrowed dramatically. That era of strong progress toward education equity preceded not only NCLB but most state high-stakes testing policies.



In the Classroom: Overtesting, Curricular Narrowing, Teaching to the Test, Cheating and Other Forms of Corruption

NCLB demanded results in the form of test data, though the bottom-line results have fallen short. The law succeeded, however, at transforming many schools into highly focused, “data-driven” environments. Testing and test preparation have proliferated—the amount of time spent on testing in some schools has doubled. A study for Congress by the Government Accountability Office (GAO) estimated states would have to create more than 433 tests (at a cost of \$1.9 billion to \$5.3 billion between 2002 and 2008) to satisfy NCLB mandates (GAO, 2003). This has become just the tip of the iceberg of a massive increase in testing. It is not uncommon for 20 to 60 school days per year to be spent in test-preparation, on top of the days spent on testing itself, which are considerable. In Massachusetts, for example, there will be 33 state test sessions across all grades this year (DESE, 2011). While the benefits of this transformation are scant, the educational costs are extremely high.

One cost is the disruption of instructional time for students who need it most. The Wisconsin Association for Supervision and Curriculum Development attempted to quantify the learning time lost to testing in general and for students with special

NAEP “provides no evidence that states can meet the laudable goal of convergence of student-subgroup performance at a significantly higher level of academic proficiency than is currently observed.”

—Aaron Pallas

Testing and test preparation have proliferated—the amount of time spent on testing in some schools has doubled.

needs. They found Wisconsin teachers spent a per-district average of 976 hours administering tests. This was particularly damaging to special needs students:

“Some schools reported that disadvantaged student populations experienced as many as 15 days—three weeks—of disrupted instructional services because the specialists were involved in test administration. Across a student’s 12-year span in a district, that could result in as many as 36 weeks, or a full year, of disrupted services for the disadvantaged students who are at the greatest risk of not meeting NCLB objectives” (Zellmer, *et al.*, 2006).

Sometimes it takes an outright scandal for NCLB-induced learning losses to come to light. At Dallas, Texas’s Field Elementary School, students were assigned grades in subjects they were not even taught in order to hide the school’s exclusive focus on NCLB’s tested subjects. This earned the school an “exemplary” rating from the state, but students were getting an education full of holes. A report from the Dallas Independent School District’s Office of Professional Responsibility (OPR) included testimony from Field teachers, who were directed by the principal to set aside music, art, and science instruction. In an email explaining why one math/science teacher should focus on math, the principal wrote, “Since the kids are so low in math, [the teacher] has to stick with math....This is a very high-stakes year and we cannot afford to have students’ TAKS scores drop in third grade” (OPR, 2011). Teachers testified that they argued against these directives, but felt they would lose their jobs if they did not comply.

Then there is the cheating epidemic that has erupted across the nation. In Atlanta, where cheating was confirmed in 44 public schools, involving 178 teachers and principals, a Georgia Bureau of Investigation (GBI) report described a culture of “fear, intimidation and retaliation spread throughout the district” (GBI, 2011). As 2011 came to a close, Georgia investigators released another report documenting widespread cheating on tests in Dougherty County, 200 miles south of Atlanta. They found evidence of cheating in each of the county’s 11 schools and similar evidence of teachers coerced into correcting students’ wrong answers. The report cited three main causes of the cheating. Reason number one: “Pressure to meet adequate yearly progress under the No Child Left Behind Act.”

Such stories of corruption and cheating in the NCLB era are so common that they cannot be dismissed as the actions of a few individuals. Instead, they are a predictable, inevitable outcome of pressure to meet test score targets, regardless of circumstances. According to published reports, incidents of cheating in the past three years have been confirmed in 30 states and the District of Columbia, whose former Chancellor Michelle Rhee has taken her “boost the scores” campaign national.

This cheating epidemic and other forms of corruption are classic examples of Campbell’s law (1976), which states, “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

“Some schools reported that disadvantaged student populations experienced as many as 15 days—three weeks—of disrupted instructional services because the specialists were involved in test administration.”

—Zellmer, et al.

Students were assigned grades in subjects they were not even taught in order to hide the school’s exclusive focus on NCLB’s tested subjects.

The report cited three main causes of the cheating. Reason number one: “Pressure to meet adequate yearly progress under the No Child Left Behind Act.”

The National Research Council (NRC) of the National Academy of Sciences looked at the accumulated evidence on test-based policies, including the federal No Child Left Behind law, state graduation tests, and policies that give teachers bonuses if their students' scores go up (Hout & Elliott, 2011). The report concluded that test-based incentives like those in NCLB increase teaching to the test and produce an inflated and inaccurate picture of what students know. It also found that educators facing sanctions tend to focus on actions that improve test scores, such as teaching test-taking strategies or drilling students closest to meeting proficiency cutoffs, rather than improving learning.

There is copious evidence of NCLB's narrowing effects from a range of sources (see, for example, Au, 2007; McMurrer, 2007; NASBE, 2003; NCES, 2007). Common Core (2011) released preliminary results of a teacher survey in December 2011. It found that 66% of teachers said NCLB's focus on math and reading has meant reduced time for art, science, and social studies. Other reports have documented how many schools are cutting recess in order to expand test preparation time, even for young children.

"During the past decade, our public schools have focused—almost exclusively—on reading and math instruction" under No Child Left Behind, said Lynn Munson, president and executive director of Common Core. Though NCLB "clearly identifies our 'core curriculum' as reading, math, science, social studies, and even the arts," many subjects have been "abandoned," Munson explained. "As a result, we are denying our students the complete education they deserve and the law demands" (Common Core, 2011).

Most troubling is that the law has exacerbated inequities it promised to end. A report from the Council for Basic Education (Von Zastrow & Janc, 2004) found evidence that narrowing was most severe in schools with higher numbers of minority and low-income students.

Linda Perlstein explained what this looks like in her book *Tested: One American School Struggles to Make the Grade* (2007). Perlstein spent a year at Tyler Elementary, a low-income school in Anne Arundel (Maryland) County school district:

That children from well-off families and children from poor ones have divergent school experiences is nothing new. What is significant is that the disparity continues in spite of (and in some ways because of) a movement designed to stop it. The practice of focusing on the tested subjects of reading and math at the expense of a well-rounded curriculum is far more prevalent where children are poor and minority. "You're not going to be a scientist if you can't read," a superintendent once told me in defense of a school's pared-down curriculum. Well, you

Test-based incentives like those in NCLB increase teaching to the test and produce an inflated and inaccurate picture of what students know.

"We are denying our students the complete education they deserve and the law demands."
—Common Core

The practice of focusing on the tested subjects of reading and math at the expense of a well-rounded curriculum is far more prevalent where children are poor and minority.

can't be a scientist—one of the most common career goals of Tyler Heights' graduating fifth-graders—if you never learn science either (p. 135).

NCLB's Role in Student Pushouts and the Growing School-to-Prison Pipeline

As bad as NCLB's narrowing and trivializing effects is the pushing out of low-scoring students to improve a school's test score bottom line. Sharon Nichols and David Berliner compiled substantial evidence of this in their book *Collateral Damage* (2007), including in Birmingham, AL, where 500 students were dropped from high school before test time, and New York City, where a lawsuit exposed policies that pushed out thousands of low-scoring students. These practices, which disproportionately affect students of color and students with disabilities, are linked to the rapid growth of a "School-to-Prison Pipeline," which is driving more and more students into the criminal justice system. The swelling of the pipeline has more than one cause, to be sure, but a 2011 position paper produced by several civil rights and education groups explained the role played by the federal testing mandate. "NCLB had the effect of encouraging low-performing schools to meet benchmarks by narrowing curriculum and instruction and de-prioritizing the educational opportunities of many students. Indeed, No Child Left Behind's 'get-tough' approach to accountability has led to more students being left even further behind, thus feeding the dropout crisis and the School-to-Prison Pipeline" (Advancement Project, et al., 2011, p. 1).

Too Much Blame, Too Little Support for Improvement

Some of NCLB's flaws might be forgiven if they had led to sustainable improvements for many schools and students in need. Instead, the law's flawed approach to accountability laid the foundation for an equally flawed and ineffective approach to providing options for parents and improving schools. A major piece of this was the provision allowing parents to transfer their children out of schools not making AYP into district schools that are. A December 2004 GAO report found fewer than 1% of the students eligible to transfer under the law did so in the 2003-04 school year. A second NCLB remedy, the supplemental services provision, has funneled money to private tutoring businesses with no measurable positive effect on students. "NCLB's Supplemental Educational Services: Is This What Our Students Need?" reported that NCLB's supplemental education services were reaching just 233,000, or 11%, of the two million students eligible nationwide, frequently offering low-quality services that merely extend NCLB's "narrowed educational agenda into students' out-of-school hours" (Ascher, 2011, p. 136).

NCLB does not invest in building new schools in failing districts, nor does it make wealthy, higher performing districts open their doors to students from poor districts. Instead, it created a menu of restructuring options for schools that fail to make Adequate Yearly Progress for six consecutive years. Such schools are subject to one of the following: takeover of the school by the state; turning management of the school over to a private firm; shutting down and reopening as a charter school; or

No Child Left Behind's 'get-tough' approach to accountability has led to more students being left even further behind, thus feeding the dropout crisis and the School-to-Prison Pipeline."

The law's flawed approach to accountability laid the foundation for an equally flawed and ineffective approach to providing options for parents and improving schools.

reconstitution of the school by replacing some or all administrators, staff, or faculty. A fifth option provided under the law endorses “any other major restructuring of a school’s governance arrangement.”

Researcher William J. Mathis looked at the record for these types of interventions in a 2009 brief. Overall, he found that there was not much of a track record for any of these approaches being used to restructure failing schools. When they were used, there was little evidence of success. For example, charter schools are rarely selected as a restructuring option, and, in any case, the record shows that “when controlling for demographic factors, charter schools show no advantage.”

Mathis (2009, p. 17) concluded: “Given that these approaches are being proposed for the nation’s most troubled schools, the solutions [currently set forth by NCLB] are likely to be woefully inadequate.” What’s more, states have no capacity to implement such sweeping restructuring remedies. While some NCLB proponents thought that the law would force states to reallocate or raise new funds to assist low-income, low-scoring schools, in general this has not been the case. Many schools remain seriously underfunded, and great funding inequalities exist both within districts and between districts in a state (FEA, 2011).

Focus on Testing Avoids Addressing the Consequences of Child Poverty

One reason why NCLB was doomed to fall short of its lofty goals has little to do with its flawed provisions or implementation. Right in the middle of the path to “100% proficiency” came the worst economic crisis since the Great Depression. According to a recent U.S. Census Bureau (2011) report, child poverty has risen to 22%, with 96 of the largest 100 school districts reporting growth in the number of poor children. Meanwhile, both school resources and the social supports children need to learn and succeed in school (housing, family and community stability, medical and dental care) are shrinking. Thus, schools must educate more and more children for whom the foundations of school success are crumbling.

The demands for equal outcomes in an unequal society would have been a dangerous illusion even without the fiscal crisis. To expect schools to counter the far-reaching impact of child poverty, to expect schools to not only keep these children from falling further behind but to accelerate their academic growth in order to close gaps in achievement, as NCLB does, is to deny reams of evidence of how poverty affects children’s ability to learn, going back to the landmark Coleman report (1966). Coleman could not have foreseen the last few decades’ staggering growth in income inequality and child poverty. But those constructing policy prescriptions during the past decade should have considered the many ways in which poverty influences a child’s ability to learn (Rothstein, 2004; Berliner, 2009).

The book *Whither Opportunity?* edited by Professors Greg Duncan and Richard Murnane (2011) documents this rising income and educational inequality and the

Charter schools are rarely selected as a restructuring option, and, in any case, the record shows that “when controlling for demographic factors, charter schools show no advantage.”

—William J. Mathis

Many schools remain seriously underfunded, and great funding inequalities exist both within districts and between districts in a state.

Schools must educate more and more children for whom the foundations of school success are crumbling.

ways in which they are linked. In their *Chicago Tribune* oped (Oct. 6, 2011), Duncan and Murnane explained:

Growing economic inequality contributes in a multitude of ways to a widening gulf between the educational outcomes of rich and poor children. In the early 1970s, the gap between what parents in the top and bottom quintiles spent on enrichment activities such as music lessons, travel and summer camps was approximately \$2,700 per year (in 2008 dollars). By 2005-2006, the difference had increased to \$7,500. Between birth and age 6, children from high-income families spend an average of 1,300 more hours than children from low-income families in “novel” places—other than at home or school, or in the care of another parent or a day care facility. This matters, because when children are asked to read science and social studies texts in the upper elementary school grades, background knowledge is critical to comprehension and academic success.

Advocates who call attention to the influence of poverty on educational outcomes are accused of making excuses for schools’ failure to close achievement gaps. The “No Excuses” proponents accuse these advocates of saying poor children cannot learn.

This charge is a red herring. A great deal can and should be done to improve schools. However, NCLB failed to consider the consequences of poverty and has been an excuse for not addressing them. Indeed, the educational “reforms” advanced in the law have predictably failed to improve schools or learning (Neill, Guisbond & Schaeffer, 2004).

On this tenth anniversary, there is ample evidence that NCLB’s false premises—that high-stakes testing coupled with sanctions would improve outcomes, without having to address other educational issues or issues of poverty—have caused the law to fail. Clearly, it is past time for a major change of course. The problem is that most proposals on the table are more of a change in rhetoric than a change in substance.

“Growing economic inequality contributes in a multitude of ways to a widening gulf between the educational outcomes of rich and poor children.”

—Duncan & Murnane

NCLB failed to consider the consequences of poverty and has been an excuse for not addressing them. Indeed, the educational “reforms” advanced in the law have predictably failed to improve schools or learning.

Part II. Sour Wine with New Labels

Secretary Duncan has clearly heard a chorus of complaints about NCLB's ill effects—from superintendents, principals, teachers, parents, students, community activists and researchers. As a result, he often speaks of flawed or limited tests and narrowed curriculum, echoing these pervasive complaints. Yet his proposals and initiatives continue most of the worst aspects of NCLB and add new, equally unsupported and harmful ones.

- Neither the Administration's waivers nor a bill from the Senate Health, Education, Labor and Pensions (HELP) Committee propose to reduce the massive overuse of standardized testing that has followed in the wake of NCLB.
- Both proposals would abandon the destructive and unrealistic AYP provision, but both keep or even expand too much of what has not worked. Duncan's Race to the Top program and waiver proposals, in particular, show the Administration's failure to consider the evidence that explains why NCLB failed.

Role of Testing Grows, Unabated

Beyond its basic testing mandates, NCLB begot a seemingly endless proliferation of tests and ways to use them: standardized tests in more subjects, interim and benchmark tests. It spawned so-called "formative" tests, which are supposed to help improve instruction but mostly take more time away from it. NCLB also fed the growth of a hugely profitable testing industry, increasing its bottom line while student achievement on NAEP leveled off and achievement gaps stagnated. Sec. Duncan heard many calls for relief from the testing avalanche and at times seemed ready to answer them.

For example, the U.S. Department of Education claimed one motivation for the waiver program was that: "NCLB has put too much emphasis on a single standardized test on a single day. This is teachers' biggest complaint about the law. They feel pressure to prepare students for those tests, leading to an unintended narrowing of the curriculum and an emphasis on the basic skills measured by standardized tests" (USDOE, 2011).

Given that rationale, the obvious response of the Administration would be to relieve this pressure. In one sense, it appears to do this, by eliminating NCLB's detested Adequate Yearly Progress mechanism. However, the dominant role of testing remains firmly in place and will likely intensify in most states.

Under the waiver plan, states must continue annual testing in reading and math of all children in grades 3-8, and once in high school, but with new tests based on "college and career standards." The evidence thus far is that the new tests will largely resemble current tests—but be harder to pass. The waivers also require states

Secretary Duncan's proposals and initiatives continue most of the worst aspects of NCLB and add new, equally unsupported and harmful ones.

Beyond its basic testing mandates, NCLB begot a seemingly endless proliferation of tests and ways to use them.

The evidence thus far is that the new tests will largely resemble current tests—but be harder to pass.

to adopt “student growth” measures and make them a “significant factor” in teacher and principal evaluation. This has pushed states to adopt statistical techniques that research shows are grossly inaccurate. Even worse, it intensifies the focus on boosting test scores instead of ensuring the all-around education of the whole child. In other words, it perpetuates the false notion that you can fatten the pig by weighing it more frequently.

The administration says that for subjects in which a state does not have tests, in order to measure “growth” there will need to be, if not more tests, then “measures that are comparable” within a district. This could push districts to buy or create dozens of new exams, at great expense and likely great damage to now-untested subjects. Charlotte-Mecklenburg, NC, for example, allocated \$1.9 million to create 52 new tests for teacher evaluation (Grundy & Sawyer, 2011; *The Herald Weekly*, 2011).

The waiver plan is particularly dangerous because most states (39 plus the District of Columbia) either have applied or say they will apply for a waiver (McNeil, 2011). Moreover, it appears unlikely that the Elementary and Secondary Education Act (ESEA), now labeled NCLB, will be reauthorized before the 2012 elections. The waiver plan could thus dramatically shape schooling in the U.S. and tighten the grip of testing (Foley & Neill, 2011).

Some states, such as California, have rejected the waivers, for reasons ranging from potentially harmful effects on education to the high costs states and communities will face to implement the waivers even as the fiscal crisis is forcing them to cut their budgets. Perhaps prior to submitting their waiver proposals, more states will decide it is a bad deal, and put pressure on Sec. Duncan to waive the AYP requirements without the quid pro quo, or poison pill, of using student test scores to judge teachers and other unwarranted waiver schemes.

The Senate HELP bill also failed to scale back the central role of testing in a reauthorized ESEA. Its bill maintains all NCLB testing; functionally defines “achievement” as test scores; and uses scores as the near-sole basis for many educational decisions (Foley & Neill, 2011). But there are additional ways in which so-called reforms ignore existing evidence and threaten to exacerbate the damaging role of testing in school.

A New High-Stakes Tool: Linking Student Test Results to Teacher Evaluations

Despite multiple studies demonstrating that linking student test scores to teacher evaluations is unfair, inaccurate and not ready for prime time, Sec. Duncan decided to carry this controversial requirement forward from Race to the Top (RTTT)² to

The administration says that for subjects in which a state does not have tests, in order to measure “growth” there will need to be, if not more tests, then “measures that are comparable” within a district.

The waiver plan could thus dramatically shape schooling in the U.S. and tighten the grip of testing.

The Senate HELP bill also failed to scale back the central role of testing in a reauthorized ESEA.

² RTTT offered competitive grants to states that agreed to comply with the administration’s favored policies, such as charter expansion, national standards and teacher evaluations linked to student test scores.

the NCLB waiver program. Once again, rather than solve the widely acknowledged problems of teaching to the test, narrowing the curriculum and perpetuating cheating and other types of corruption, this “innovation” will exacerbate them by making test results even more high stakes for teachers.

In November 2011, the Education Writers Association released a brief on teacher evaluation based on more than 40 studies and interviews with scholars (Sawchuk, 2011). The brief concludes that existing research does not support linking teacher evaluations to student test scores, for multiple reasons:

- Teachers are not the most important factor in student achievement, which is mostly a product of individual and family background characteristics.
- The politically popular value-added methods of measuring teachers are generally not reliable or stable. These measures may pick up some differences in teacher quality, but they can be influenced by a number of factors, including statistical controls and characteristics of schools and peers.
- Contrary to claims that student achievement can be greatly influenced by having highly effective teachers several years in a row, a teacher’s effectiveness varies from year to year. The impact of an effective teacher seems to decrease with time, so the cumulative effects of having better teachers for several years in a row are not clear.
- In the United States, rewarding teachers whose students produce gains (sometimes termed “merit pay”) has not been shown to improve student achievement.

University of California-Berkeley economist Jesse Rothstein (2010) offered the most succinct assessment of value-added methods, concluding they are “only slightly better than coin tosses” at measuring teacher quality.

Despite the evidence (see also Neill, 2011), because of pressure and incentives from RTTT and the NCLB waiver program, states have aggressively moved forward in planning and implementing teacher evaluation programs linked to student test results. At least 23 states and Washington, D.C. evaluate teachers in part by test scores, and 14 states allow districts to use data to dismiss teachers.

Tennessee was a winner in the RTTT competition, and swiftly implemented an evaluation system that bases half of a teacher’s assessment on “student achievement,” with 35% to come from growth measures based on student scores on state tests, and requires frequent evaluations by principals. The system caused such frustration and confusion that State Education Commissioner Kevin Huffman quickly called for modifications. One problem is that, as in most states, most teachers teach untested subjects, so there are no test scores to evaluate them. Tennessee “solved” this problem by allowing teachers to be evaluated with scores in a subject they do

Existing research does not support linking teacher evaluations to student test scores, for multiple reasons.

In the United States, rewarding teachers whose students produce gains (sometimes termed “merit pay”) has not been shown to improve student achievement.

not teach. For example, a gym teacher could “choose” to be evaluated by students’ scores in writing. Will Shelton, principal of Blackman Middle School in Tennessee, described the system: “I’ve never seen such nonsense. In the five years I’ve been principal here, I’ve never known so little about what’s going on in my own building” (Winerip, 2011).

A similar New York plan has led to an unprecedented explosion of resistance from principals. As of late December, nearly 23% of New York State principals (1,058) had signed a protest statement objecting to “an unproven system that is wasteful of increasingly limited resources. More importantly, it will prove to be deeply demoralizing to educators and harmful to the children in our care. Our students are more than the sum of their test scores, and an overemphasis on test scores will *not* result in better learning” (N.Y. Principals, 2011).

Despite Lack of Success, Turnaround Strategies Stay on the Program

The administration’s waiver program and the Senate HELP bill eliminate AYP and instead create much more limited categories of schools requiring intervention. States granted waivers must focus turnaround efforts only on the lowest performing 5%, so-called “priority” schools. Another 10% (“focus” schools) would have interventions targeted at their lowest performing student groups, which could include the transfer and tutoring options that were unsuccessful in NCLB. The Senate HELP bill similarly identifies the lowest 5% of schools, based on test scores and (for high schools) graduation rates, for interventions.

This nod to states’ limited capacity was a small dose of reality. But the waiver program’s turnaround alternatives hew closely to NCLB, which included reopening as a charter school, replacing most or all of the staff, turning governance over to an outside entity, or “any other major restructuring.”

The waiver options include:

- A turnaround model that would replace the principal and rehire no more than half the school staff.
- A restart model in which the school is converted or closed and reopened under an education management organization, which could be a charter.
- The school could simply be closed and its students enrolled in other schools.
- A transformation model, “which address four areas critical to transforming persistently lowest-achieving schools. These areas include: developing teacher and principal leader effectiveness, implementing comprehensive instructional reform strategies, extending learning time and creating community connections, and providing operating flexibility and sustained support.”

The Tennessee system caused such frustration and confusion that State Education Commissioner Kevin Huffman quickly called for modifications.

Nearly 23% of New York State principals (1,058) had signed a protest statement objecting to “an unproven system that is wasteful of increasingly limited resources. More importantly, it will prove to be deeply demoralizing to educators and harmful to the children in our care.”

The latter model is a softer version of RTTT requirements and at least offers the possibility of taking genuine improvement steps. No federal funding is provided, however, for these costly school improvement approaches.

The Senate HELP bill's turnaround strategies are also similar, though the list includes an additional three options that could allow more flexibility (Foley & Neill, 2011). In addition, HELP requires districts to develop specific improvement actions for all its "5%" schools, based on a review of the institution and rooted in such things as professional development and collaboration that have proven to lead to genuine school improvement when done well. To some meaningful degree, here the Committee responded to recommendations from education and other organizations (FEA, 2011b).

Without evidence, administration officials turn to anecdote, highlighting "model" schools to show that their prescriptions have worked in the past. Education scholar Diane Ravitch, however, looked closely and found no miracles (Ravitch, 2011b). For example, the president traveled to Florida in March to join Gov. Jeb Bush in praising Miami Central High for its transformation, after more than half the staff had been fired. Ravitch found that this "miracle school" remains one of Florida's lowest performers and narrowly evaded closure.

While miracle turnarounds are rare, evidence-based strategies for improving school performance have shown success. This report's next section briefly summarizes alternative approaches to NCLB reform that include these real-world approaches.

The waiver program's turnaround alternatives hew closely to NCLB, which included reopening as a charter school, replacing most or all of the staff, turning governance over to an outside entity, or "any other major restructuring."

HELP requires districts to develop specific improvement actions for all its "5%" schools, based on a review of the institution and rooted in such things as professional development and collaboration that have proven to lead to genuine school improvement when done well.

Part III. Real Reform Is Possible but Would Mean Setting a New, Evidence-Based Direction

NCLB's authors tried to give the law a level of gravitas by calling for "scientifically based research" (SBR) to guide educational practice. The law defined SBR as "research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs." Many, including FairTest (Neill, Guisbond & Schaeffer, 2004), argued from the start that existing research on high-stakes testing, turnaround strategies and other aspects of the law should have led NCLB's authors in very different directions. Unfortunately, these pleas had little effect.

Fortunately, there are models of successful practice here and abroad that could form the basis for a revised ESEA, which would support meaningful, sustainable educational reforms.

Linda Darling-Hammond (2010), Diane Ravitch (2011a), Tony Wagner (2011) and many others have observed that top-ranked Finland, for example, dramatically reformed its schools over 20 years to achieve a high degree of equity and quality. It did so by pursuing policies that diametrically oppose those in NCLB. Finland has no high-stakes testing to rank students or schools and does not evaluate teachers based on student test scores. There are no mass firings of teachers, closures of "failing schools" or turnaround experts brought in to shake things up. There are no scripted curricula or frequent benchmark assessments leading up to a big test.

Instead, Finland focuses on ensuring equitable educational resources across the board – even providing more to schools serving students with the greatest needs. It has developed a strong, unionized teaching force that works together to improve schooling, and a sound, comprehensive curriculum. Finland invests heavily in teacher preparation and development and then gives their well-prepared and supported teachers tremendous autonomy and respect. Other top-ranked nations such as Singapore and Hong Kong pursue similar approaches, and their students reap the benefits (Darling-Hammond, 2011).

It is more difficult to find successful comprehensive models in the U.S., in part because of a profoundly unequal society and in part because all public schools are ruled by NCLB's rigid requirements. But on a smaller scale, there are schools that demonstrate better methods of assessment, such as the use of multiple measures of student learning (FairTest, 2010a). For example, the New York Performance Standards Consortium (n.d.) high schools have taken advantage of a variance from the New York Board of Regents to limit standardized testing to just the state English Language Arts (ELA) exam. The schools fulfill state and federal accountability requirements using the ELA test along with their own math and language arts tasks and other performance-based assessments.

Fortunately, there are models of successful practice here and abroad that could form the basis for a revised ESEA, which would support meaningful, sustainable educational reforms.

Finland has no high-stakes testing to rank students or schools and does not evaluate teachers based on student test scores.

Finland invests heavily in teacher preparation and development and then gives their well-prepared and supported teachers tremendous autonomy and respect.

This flexibility has allowed these schools, whose demographic makeup is roughly comparable to New York City public schools, to create a system with extremely high expectations and outcomes. According to the Consortium's web site (n.d.), "The [performance] tasks require students to demonstrate accomplishment in analytic thinking, reading comprehension, research writing skills, the application of mathematical computation and problem-solving skills, computer technology, the utilization of the scientific method in undertaking science research, appreciation of and performance skills in the arts, service learning and school to career skills." Dropout rates are very low (9.9% compared to the citywide rate of 19.3%) and college acceptance rates high (91% compared with the citywide rate of 62.6%). It is easy to imagine how much more they and others could accomplish within the context of a federal law that actively encouraged and supported such an approach.

Alternative Proposals for NCLB Reform

More than 150 national education, civil rights, religious, disability, civic, labor and other groups have now signed the *Joint Organizational Statement on NCLB*. The statement enumerated problems with NCLB and recommended reforms, including a move away from the "overwhelming reliance on standardized tests to using multiple indicators of student achievement in addition to these tests."

Out of that initiative came the Forum on Educational Accountability (FEA), which in 2011 laid out a proposal for NCLB reauthorization (FEA, 2011b). The plan would ensure that schools have the capacity to help all children achieve success while outlining a reasonable federal role in educational policy instead of top-down mandates that are too often overly prescriptive and fail to help schools reach desired educational and societal goals. FEA's proposals cover four main areas: overhauling assessment, restructuring accountability, developing school capacity to serve all students well, and addressing the unmet human and social needs faced by many children. Each is rooted in solid evidence—research and practical experience from the U.S. and other nations.

Relying on work done in Massachusetts, FairTest (2010b) has promoted a three-part assessment and evaluation program each state could implement. The plan includes gathering and evaluating classroom- and school-based evidence of student learning each year for each school; administering low-stakes standardized statewide tests in reading and math to each student every few years; and using "school quality reviews" which involve teams of experts conducting a careful review of each school every 4-6 years to ascertain how well it is meeting the full range of student needs. Together, this evidence would provide a far richer picture of student learning and school progress than can standardized tests alone. At the same time, it would avoid the damaging consequences of NCLB such as teaching to the test and narrowing the curriculum.

"The [performance] tasks require students to demonstrate accomplishment in analytic thinking, reading comprehension, research writing skills, the application of mathematical computation and problem-solving skills, computer technology, the utilization of the scientific method in undertaking science research, appreciation of and performance skills in the arts, service learning, and school to career skills."

—New York Performance Standards Consortium

FEA's proposals cover four main areas: overhauling assessment, restructuring accountability, developing school capacity to serve all students well, and addressing the unmet human and social needs faced by many children.

Proposals from other groups, such as the Broader, Bolder Agenda for Education (2008), share fundamental characteristics with FairTest and FEA reform proposals that sharply distinguish them from NCLB and its offspring. These alternatives explicitly acknowledge the need for dramatic and fundamental, not incremental, changes to the law. They recognize that the purpose of assessment is to help teachers improve instruction and to strengthen schools, not to label and punish them. They support methods to identify needed improvements more precisely in order to better target school efforts and outside assistance. They recognize the need for educational and other public policies to address health, social, emotional and other basic needs. They call for the education of the whole child and for gathering a range of data relevant to that goal, including survey data on school climate.

NCLB was promoted as an example of a bipartisan consensus on education policy. Over time, it has become clear that the details of the law, if not some of its stated goals, were fundamentally flawed. This report has summarized the evidence of what went wrong as well as the nation's ongoing education policy challenge and how to confront it in order to meaningfully improve public education. After a decade of stagnation, it is now time to use this evidence to craft and pass a new federal education law that will help and not harm our schoolchildren.

This evidence would provide a far richer picture of student learning and school progress than can standardized tests alone.

After a decade of stagnation, it is now time to use this evidence to craft and pass a new federal education law that will help and not harm our schoolchildren.

References

- Advancement Project, *et al.* 2011, March (revised). Available online at <http://www.scribd.com/doc/51122359/Advancement-Project-School-to-Prison-Pipeline-Position-Paper>
- Ascher, C. 2006. NCLB's Supplemental Educational Services: Is This What Our Students Need? *Phi Delta Kappan*, Vol. 88, No. 02: pages 136-141.
- Au, W. 2007. *High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis*. *Educational Researcher*, Vol. 36, No. 5, pp. 258–267
- Berliner, D. 2009. *Poverty and Potential: Out of School Factors and School Success*. Available online at <http://nepc.colorado.edu/publication/poverty-and-potential>
- Broader, Bolder Approach to Education. 2008. Available online at <http://www.bolderapproach.org/index.php?id=51>
- Campbell, D. T. 1976. *Assessing the Impact of Planned Social Change*. The Public Affairs Center, Dartmouth College.
- Coleman, J. et al. 1966. *Equality of Educational Opportunity*. Government Printing Office
- Common Core/Farkas Duffett Research Group. 2011. Learning Less: Public School Teachers Describe a Narrowing Curriculum. Highlights from a survey by Common Core and the Farkas Duffett Research Group. Available online at <http://www.commoncore.org>
- Darling-Hammond, L. 2010. *The Flat World and Education: How America's Commitment to Equity Will Determine Our Future*. New York: Teachers College Press, Teachers College, Columbia University.
- Duncan, G. & Murnane, R. 2011. *Whither Opportunity? Rising Inequality, Schools and Children's Life Chances*. New York: Russell Sage.
- Duncan, G. & Murnane, R. Oct. 6, 2011. "Economic inequality: The real cause of the urban school problem." *The Chicago Tribune*. Available online at http://articles.chicagotribune.com/2011-10-06/news/ct-perspec-1006-urban-20111006_1_poor-children-graduation-rate-gap
- FairTest. 2010a. Multiple Measures: A Definition and Examples from the U.S. and Other Nations. Available online at <http://www.fairtest.org/fact-sheet-multiple-measures-definition-and-exampl>

FairTest. 2010b. A Better Way to Evaluate Schools. Available online at <http://www.fairtest.org/fact-sheet-better-way-evaluate-schools-pdf>

FairTest. 2009. NAEP Results Produce More Evidence of NCLB's Failure. Available online at <http://www.fairtest.org/naep-results-produce-more-evidence-nclbs-failure>

Foley, B. & Neill, M. 2011. *Is AYP Really Dead? A Look at Waivers and ESEA Reauthorization*. Available online at <http://fairtest.org/slide-show-basic-facts-about-duncans-ayp-waiver-pl>

Forum on Educational Accountability (FEA). 2011a. *All Children Deserve the Opportunity to Learn*. Available online at <http://www.edaccountability.org/reports.html>

Forum on Educational Accountability (FEA). 2011b. FEA Recommendations for Improving ESEA/NCLB. Available online at <http://www.edaccountability.org/Legislative.html>

Georgia Bureau of Investigation (GBI). 2011. *Special Investigation into CRCT Cheating at APS*. Available online at <http://www.ajc.com/news/volume-1-of-special-1000798.html>

Georgia Bureau of Investigation (GBI). 2011b. *Dougherty County Investigative Report, Vols. 1 and 2*. Available online at <http://www.gaosa.org/>

Grundy, P. & Sawyer, C. May 15, 2011. The push behind a new flurry of testing. *The News Observer*. Available online at <http://www.newsobserver.com/2011/05/15/1196022/the-push-behind-a-new-flurry-of.html>

The Herald Weekly. 2011. School district to collaborate with state on test development. Available online at <http://www.huntersvilleherald.com/education/2011/12/22/school-district-to-collaborate-with-state-on-test-development/>

Hout, M. & Elliott, S. Editors. 2011. *Incentives and Test-Based Accountability in Education*. Committee on Incentives and Test-Based Accountability in Public Education; National Research Council. Available online at http://www.nap.edu/catalog.php?record_id=12521

Joint Organizational Statement on No Child Left Behind Act. 2004. Available online at <http://www.edaccountability.org>

Massachusetts Department of Elementary and Secondary Education (DESE). 2011. Massachusetts Comprehensive Assessment System Schedule for MCAS and MEPA Testing. Available online at <http://www.doe.mass.edu/mcas/cal.html>

Mathis, W.J. 2009. *NCLB's Ultimate Restructuring Alternatives: Do They Improve the Quality of Education?* Boulder and Tempe: Education and the Public Interest Research Center & Education Policy Research Unit. Available online at <http://nepc.colorado.edu/publication/nclb-ultimate-restructuring>

McMurrer, J. 2007. *NCLB Year 5: Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era*, Center on Education Policy. Available online at <http://www.cep-dc.org/displayDocument.cfm?DocumentID=312>

National Association of State Boards of Education (NASBE). 2003. *The complete curriculum. Ensuring a place for the arts and foreign languages in America's schools. The report of the NASBE study group on the lost curriculum*. Alexandria, VA

National Center for Educational Statistics (NCES). 2011a. *The Nation's Report Card: Mathematics 2011*. Available online at <http://nces.ed.gov/nationsreportcard/>

National Center for Educational Statistics (NCES). 2011b. *The Nation's Report Card: Reading 2011*. Available online at <http://nces.ed.gov/nationsreportcard/>

National Center for Education Statistics (NCES). 2008. *The Nation's Report Card: Long-Term Trend 2008*. Available online at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009479>

National Center for Education Statistics (NCES). 2007. *Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 Through 4* (Issue Brief). Available online at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007305>

McNeil, M. 2011. Majority of States Say They'll Seek Waivers Under NCLB. *Education Week*. Available online at <http://www.edweek.org/ew/articles/2011/10/13/08waivers.h31.html?r=953279945>

Neill, M. 2011. Student Test Scores: An Inaccurate Way to Judge Teachers. Available online at http://fairtest.org/sites/default/files/Student_Test_Scores_dangerous_way_to_evaluate_teachers_-_4-24-11.pdf

Neill, M., Guisbond, L., and Schaeffer, R. 2004. *Failing Our Children: How "No Child Left Behind" Undermines Quality and Equity in Education; An Accountability Model that Supports School Improvement*. Available online at <http://fairtest.org/node/1778>

New York Performance Standards Consortium. N.d. <http://www.performanceassessment.org>

New York Principals. 2011. An Open Letter of Concern Regarding New York State's APPR Legislation for the Evaluation of Teachers and Principals. Available online at <http://www.newyorkprincipals.org/>

Nichols, S. & Berliner, D. 2007. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge, MA: Harvard Education Press.
Office of Professional Responsibility. 2011. *Report of Investigation*. Available online at <http://www.scribd.com/doc/73055392/DISD-Field-Elementary-investigation>

Pallas, A. 2011. The Nation's Report Card and NCLB: Friends or Foes? National Education Policy Center. Available online at <http://nepc.colorado.edu/blog/nation%E2%80%99s-report-card-and-nclb-friends-or-foes>

Perlstein, L. 2007. *Tested: One American School Struggles to Make the Grade*. New York: Henry Holt.

Ravitch, D. 2011a, Oct 11. What Can We Learn from Finland? Available online at http://blogs.edweek.org/edweek/Bridging-Differences/2011/10/what_can_we_learn_from_finland.html

Ravitch, D. 2011b, Dec. 6. Do You Believe in Miracles? Available online at http://blogs.edweek.org/edweek/Bridging-Differences/2011/12/do_you_believe_in_miracles.html

Rothstein, J. 2011. Review of "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project." Boulder, CO: National Education Policy Center. Available online at <http://nepc.colorado.edu/thinktank/review-learning-about-teaching>

Rothstein, R. et al. 2008. *Narrowing the Achievement Gap for Low-Income Children: A 19-Year Life Cycle Approach*. Washington, D.C.: Economic Policy Institute. Available online at http://www.epi.org/publication/narrowing_the_achievement_gap_for_low-income_children/

Rothstein, R. 2004. *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington, D.C.: Economic Policy Institute. Available online at http://www.epi.org/publication/books_class_and_schools/

Sawchuk, S. Nov. 12, 2011. *What Studies Say about Teacher Effectiveness*. Education Writers Association. Available online at http://www.ewa.org/site/PageServer?pagename=research_teacher_effectiveness

U.S. Census Bureau. 2011. *Income, Poverty and Health Insurance Coverage in the United States: 2010*. Available online at http://www.census.gov/newsroom/releases/archives/income_wealth/cb11-157.html.

U.S. Department of Education (DOE). 2011. The New Consensus on Middle-Grades Reform. (Remarks of U.S. Secretary of Education Arne Duncan to the Association for Middle Level Education (AMLE) Annual Conference.) Available online at <http://www.ed.gov/news/speeches/new-consensus-middle-grades-reform>

U.S. Government Accountability Office (GAO). 2003. *Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies*. Available online at www.gao.gov/new.items/d03389.pdf

U.S. Government Accountability Office (GAO). 2004. *Education Needs to Provide Additional Technical Assistance and Conduct Implementation Studies for School Choice Provision*. Available online at <http://www.gao.gov/products/GAO-05-7>

Wagner, T. 2011. *The Finland Phenomenon: Inside the World's Most Surprising School System*. <http://www.2mminutes.com/products/pc/viewPrd.asp?idProduct=22&idcategory=24>

Winerip, M. 2011. In Tennessee, Following the Rules for Evaluations Off a Cliff. *The New York Times*. Available online at <http://www.nytimes.com/2011/11/07/education/tennessees-rules-on-teacher-evaluations-bring-frustration.html?pagewanted=all>

Von Zastrow, C. & Janc, H. 2004. Academic Atrophy: The Condition of Liberal Arts in America's Public Schools. Council for Basic Education. Available online at <http://www.ecs.org/html/Document.asp?chouseid=5058>

Zellmer, M., *et al.* 2006. What Are NCLB's Instructional Costs? *Educational Leadership*. Association for Supervision and Curriculum Development (ASCD). Available online at <http://www.ascd.org/publications/educational-leadership/nov06/vol64/num03/What-Are-NCLB%27s-Instructional-Costs%C2%A2.aspx>