

REAL AND ILLUSORY GAINS IN PROFICIENCY UNDER NO CHILD LEFT BEHIND

THOMAS GONZALEZ, UNIVERSITY OF CHICAGO

RESEARCH GUIDED BY:

AUSTIN NICHOLS, DOUG WISSOKER, AND COLIN TAYLOR

OVERVIEW

In 2001, Congress reauthorized the Elementary and Secondary Education Act of 1965 in the form of the bipartisan No Child Left Behind Act (NCLB). NCLB created a mandate for states to close achievement gaps and position public school systems to meet the challenges of the 21st century. However, the absence of specific national standards under NCLB combined with punitive consequences for poor performance creates incentives for states to develop lenient standards. Some states are responding to greater accountability under NCLB by maintaining low standards and some are adopting lower standards over time to create the appearance of improvement. Despite perverse incentives, other states find ways to improve student performance while adhering to high standards.

The lack of uniform standards makes it difficult for policy makers and educators to identify states that are truly improving the quality of their states' education system. Without the information provided by uniform standards, policy makers and educators cannot deliver financial and technical support to those schools that need help, nor can they apply proven strategies to low-performing states.

I find that stricter standards may actually lead to improved student outcomes on a nationally normed test, so those states with lower standards may be sacrificing their students' futures in order meet administrative targets for pass rates.

A NEED FOR REFORM

America's public education system bears the responsibility of preparing our children to compete in today's global and technologically intensive economy. Legislation must address racial and socioeconomic achievement gaps at the same time that it promotes high standards for all American public schools.

Rapidly developing international economies and advances in technological communication will force the next generation of Americans to compete for jobs in an international labor pool. While American firms move overseas to take advantage of both skilled and unskilled labor at a relatively cheaper cost than the US labor market, the American public education system has failed to keep pace. Three recent international exams, The Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and The Program for International Student Assessment (PISA) show that other

countries are catching up to and surpassing American children in key subjects. Despite relatively high investment in education and smaller class sizes, US students ranked 15th in the world on the 2003 TIMSS and 11th in science¹. The PIRLS assessment of 2006 showed no advance from 2003 scores for American 4th graders, who scored 15th on the exam down from 9th in 2001². In 2006, American 15-year-olds ranked 35th on the math portion of PISA, an exam geared toward real-world applications of mathematics³.

Troubling domestic disparities exist as well. US Census data for 2000 showed that 60 percent of non-Hispanic Whites between the ages of 18 and 19 had a high school degree or higher compared to 50 percent of non-Hispanic Blacks and 44 percent of Hispanics⁴. No Child Left Behind aims to close this racial achievement gap while raising standards for all American public schools.

NCLB emphasizes four ideas that separate it from past legislation:

- increased accountability at the local level for student performance,
- greater choice for parents in deciding which school their child attends,
- mandatory qualifications for teachers in their subject area, and
- implementation of research-based teaching methods to improve student performance.

The success of these ideas is contingent on the collection of reliable information to assess how students are performing, which schools need support, and which style of teaching yields results.

To measure performance, states test grades 3 through 8, though before 2005-2006, states could choose to test only two grades from this range⁵. Although critics of NCLB fiercely debate the merits of using standardized tests as a way to evaluate a child's education, this paper focuses primarily on how differences in testing standards across states

under NCLB lead to different measurement of student outcomes on state tests.

THE LIMITED ROLE OF THE FEDERAL GOVERNMENT IN EDUCATION

Understanding the limitations of the federal government's ability to influence state and local education policy is important in evaluating potential reforms. No Child Left Behind places the federal government in a historically unusual position of authority in shaping public education. With less than one-tenth of public school funding coming from the federal government, state expenditures comprise 47 percent of school budgets and local taxes, the remaining 43 percent⁶. Disparities in local property tax values lead to inequalities in available revenue for public schools, likely a driving factor behind racial and income achievement gaps.

Despite the apparent need for government to even the playing field, Article 1, section 8 of the Constitution does not grant Congress the power to intervene in education: the federal government is limited to affecting public education indirectly by setting conditions under which states and schools qualify for federal funds. Title 1 of NCLB allocates funds to Local Education Agencies (LEAs), which give money to schools with a large proportion of low income or minority students⁷. Federal dollars help states by lowering the costs that Title 1 schools face; however, states continue to shoulder a sizeable portion of these costs. Given the constitutional limits on federal authority, NCLB might not be deemed constitutional were it not for the autonomy granted to states in determining how to implement accountability.

ADEQUATE YEARLY PROGRESS

No Child Left Behind requires that public schools make Adequate Yearly Progress (AYP). AYP goals include having all students proficient on their

respective state exam by the year 2014, graduating a pre-determined percent of high school students*, making all schools “safe and drug free,” and staffing every public school classroom with a highly qualified teacher⁸. Though previous studies comprehensively examine these goals, my analysis focuses on the goal of 100 percent proficiency on state tests, an area where states have great flexibility. Flexibility arises because states design their own tests and can change test format and content from year to year.

Each state has a schedule for AYP growth requirements in any given year, which the law calls “annual measurable objectives” (AMOs). For example, in 2002, Missouri submitted to the Department of Education a request to implement the following AMO timeline: in 2008, for a Missouri school district to make “Adequate Yearly Progress,” 45 percent of its students would have to be proficient in math—that target would jump to 54 percent in 2009 and 63 percent in 2010⁹. If any subgroup within a public school fails to make AYP for two consecutive years, the state identifies that school for targeted interventions. Subgroups include whites, blacks, Asians, Native Americans, Hispanics, English Language Learners, students with disabilities, and low-income students.

Title 1 schools (schools with large numbers of low income or minority students) that miss AYP for two consecutive years must alert parents of the option to transfer their child to another school in that district at the expense of Title 1 funds. These schools must also devise a plan on how to improve student performance, and finally, they must allocate 10 percent of their Title 1 funds to professional development (states apply these sanctions to non-Title 1 schools at their own cost¹⁰). If that school misses AYP the following year, it must provide students with supplemental educational services

such as free tutoring or other instruction from a state-approved private provider.

Persistent failure to meet AYP standards results in more drastic measures. Schools that fail an additional year must employ outside consultants who will implement research-based methods, take authority from administrators and replace teachers. If a school in the process of restructuring fails again, it may be converted into a charter school or face complete state control¹¹. In Philadelphia and DC, school district control has shifted from local school boards to state or mayoral hands. Though expansions in Title 1 funding cover part of these expenses, most states face the pressure of tighter budgets coinciding with more schools failing to meet AYP standards¹².

States have strong fiscal incentives to avoid such remedial measures and shy away from challenging standards, despite the designation of funds for Title 1 school districts that face these costs. Delaware, Wisconsin and Arkansas report over 98 percent of school districts meeting AYP for 2004-2005 testing while fewer than 10 percent of school districts made AYP in West Virginia, Florida and North Carolina¹³. One legitimate explanation for this disparity is that half of the fifty states planned their “annual measurement objective” trajectory so that school districts faced low standards in the first few years after NCLB. These states are relying on a rapid increase in the percent of students passing in the years leading up to 2014. This approach may have allowed more time for schools to adjust to changes in curricula and new standards compared to schools in states where AMOs move at constant incremental levels¹⁴. Despite different AMO trajectories, California is one of several states to estimate that ninety nine percent of its school districts would not meet one hundred percent proficiency by the year 2014¹⁵. Whatever the story, this picture suggests a wide divergence in AYP standards.

* See discussion of dropout rates on page 5

MAKING PROGRESS OR LOWERING STANDARDS?

A recent study provides evidence that some states lower their standards to gain ground on AYP indicators (high school graduation rates, school safety, etc) and raise the percent of students scoring at the proficient level. Fordham Institute researchers examined differences in the difficulty of concepts tested and test format, and the range in proficiency cutoff points used on state exams and found great disparities across states and also within states. Reading and math standards were different and standards across grades were different. In other words, if a student passed the third grade reading test, the student might still be more than a year's worth of instruction away from passing the 4th grade reading test¹⁶.

According to the Fordham study, 4th graders in Massachusetts, the state with the most rigorous standards, are expected to read and analyze a passage of Tolstoy, while fourth graders in other states answer more elementary questions¹⁷. Massachusetts also has four open ended questions, the most among any state, whereas some states use only multiple-choice format¹⁸. The cost of grading differently formatted tests influences this decision. Also, though open-ended questions require more effort on the part of the students, they may be scored more or less stringently from year to year or from grader to grader.

The Fordham study used the Measures of Academic Progress exam (MAP), a national exam, as a yardstick for comparing proficiency standards in different states. On the 2006 8th grade math portion, students who were below the 55th percentile on the MAP test failed to score "proficient" on the Massachusetts, Montana and New Mexico state tests. Using the same technique of comparing state test scores and MAP, the Fordham Institute found that Illinois, Wisconsin, and Colorado passed

students on their state exams that were below the 25th percentile on the MAP exam¹⁹.

Many states institute a minimum group size to determine whether a subgroup qualifies towards AYP evaluation. If the state's minimum group size were set at 40 students and a school district had only 35 Hispanic students, the scores of Hispanics would not count towards AYP evaluation. Most states have smaller minimum group sizes requirements, however this loophole allows states room to "leave behind" subgroups²⁰. Some states have also adopted a rule requiring all three levels of schools in a district, elementary, middle and high school, to fail before that district fails AYP. Not only that, but they must fail in the same subject²¹. Missouri is one of the many states that recently submitted a request to add flexibility to its original growth model. This "Safe Harbor" condition of NCLB allows schools to make AYP if they reduce the percent of students that failed to make AYP the previous year by 10 percent²². To account for guessing on tests, differences in grading, and possible distractions during test takings, NCLB allows states to use a confidence interval to pass students whose scores are not statistically different from a proficient score. Some states use a 99 percent interval, which is arguably useless in that it counts people far below the proficiency mark as proficient²³. Before 2004, some states were allowing students to retake exams if they did not score proficient. Though the Department of Education had never explicitly agreed to this practice it eventually incorporated it into NCLB²⁴.

States also set their own standards for teacher qualifications, one of several other indicators for AYP. High Objective Uniform Standard of State Evaluation, the section of NCLB outlining the goal of teacher qualifications, grants discretion to states in setting standards. From 2004-2005, Wisconsin, Montana, and Oklahoma reported over 99 percent of teachers in core classes to be highly qualified while Nevada, DC and Alaska reported less than 75 percent of teachers to be qualified. Some states

require teachers to pass a test in their subject area while other states have teachers fill out a form broadly outlining their knowledge area and their experience in the field²⁵.

High school graduation rate is another AYP indicator. Here too, states set their own formulas. Some states count graduation rates by taking the ratio of students that start 12th grade and graduate one year later, thereby excluding students that dropped out before 12th grade. A more accurate measure of high school graduation rates is the ratio of students who graduate in any given year to the number of students who entered high school four years before²⁶. Furthermore, states were allowed to set their own targets for long-term graduation rates resulting in 2014 goals that range from 65 to 100 percent²⁷.

Another component of AYP is the goal of having all schools safe by 2014. Arizona, among other states, has lower standards for student safety than schools in Maryland. Arizona's evaluation of a dangerous school includes only reported firearm incidences and ignores expulsion rates, drug violations and violence, which Maryland and a few other states incorporate in their criteria²⁸.

If we look at these standard setting procedures individually, they are fair in allowing some leniency to states; however, taken together, these adjustable standards allow for wide variation in AYP standards. Avoiding higher standards undermines NCLB and prevents parents and educators from recognizing which states are performing well and moving towards the goals of NCLB, and which states are not. If all 4th graders are taking a test at the 2nd grade level and everyone scores "proficient," we cannot know if the achievement gap has closed since all subgroups passed. On the other hand, if the test is really hard and no one passes, the achievement gap has also artificially closed. Therefore, the cutoff point is important in determining real changes in the achievement gap.

ANALYSIS AND RESEARCH QUESTIONS

The following three questions structured my analysis:

1. *Did standards differ among states in 2005?*
2. *Did state standards change from 2005 to 2007?*
3. *Did higher state standards in 2005 correlate with greater improvement on a national exam?*

To answer these questions, we need to have a nationally standardized and reliable measure of student achievement levels that is comparable across states. I used the National Assessment of Educational Progress (NAEP), a nationally administered exam in use since the 1970's as a yardstick for cross-state comparison. In addition to the state-designed test, NCLB mandates that states conduct random testing of students in the 4th and 8th grade on NAEP every two years. Because of the availability of these standardized test results, I chose to examine 4th and 8th grade student performance on state tests compared to a national test (without the same punitive incentive structure). Though there are other tests such as MAP that could be used to compare states, NAEP is often reported when showing trends in achievement in the pre and post-NCLB years*. In fact, from 2003 to 2007, the percent of students passing the math section of the NAEP grew by 3 percent for 8th graders and declined by 1 percent in reading. Critics of NCLB may cite this statistic as evidence for the failure of NCLB; however, this statistic represents national trends and

* One potential problem in comparing NAEP and state test scores is that they may test different material. Poor performance on the NAEP may not reflect that students are failing to improve but that the NAEP is more aligned with the curricula of those states that perform well on it. For the purpose of this study, I assume that state and NAEP tests cover similar material and that changes in their alignment signify changes in standards.

ignores different trends across states. The important lesson to take away is that NAEP scores have never moved quickly in the last thirty years. Therefore, small growth is more impressive than it may at first seem.

To determine a state's educational testing standard I compared the percentage passing the state test with the percentage passing the NAEP test. I argue that states that maintained high standards on their NCLB tests would show identical performance on state exams and the NAEP (ie: if half of the 4th graders in Tennessee were able to pass the NAEP reading test, and the state tested at the same level for NCLB, the same 50 percent of Tennessee 4th graders should pass their state reading test). I expected, however, that due to low standards, a greater percentage of students would pass state exams than the NAEP. Ultimately, I wanted to determine whether states that maintained high standards saw more improvement on the NAEP over time.

METHODOLOGY

For the purposes of this study, I compare state and national test scores over the five-year period after the implementation of NCLB—2003 to 2007 (the most recent year for which data are available). The NAEP data set that I used for this study comes from School Data Direct (<http://www.schooldatadirect.org/>), an organization that compiles NAEP scores over time and which the USDOE references on its NCLB page. For state level data, I used each state's "State Report Card," reported by State Education Agencies. Not all states had data available for 2003 to 2007 so I filled in missing data using various sources including the Center on Education Policy's 2007 publication, *Has Student Achievement Increased Since 2002? State Test Score Trends Through 2006–07*²⁹. The available state data limited which states I could use in the analysis. I did not include DC, Alaska, Vermont, Rhode Island or New Hampshire because they had no available data in 2005 for grades 3, 4, 5, 7 or 8. Where data reported by state departments differed

from outside sources, I used the percent proficient that the state education agency reported because it captures how local constituencies perceive the performance of their state's public education system. Colorado included "Partially Proficient" in the percent of proficient students and South Carolina included "Basic" in the group, "Meets Standard."

To address my first research question—whether testing standards varied across states in 2005—I did not attempt to measure the difficulty of a state test but only the difficulty relative to the NAEP. Figure 1 illustrates this distinction. In this example, State A has "more proficient" students because more of them passed the nationally normed NAEP. However, within State A, more students passed the state test; therefore the state test was easier than the NAEP. State B's students scored lower than State A's students on the NAEP but they had a harder state test relative to the NAEP.

To determine the relative difficulty or "strictness" of each state's test, I use the difference between the percent of proficient students on the NAEP exam and the percent of proficient students on the state test for 4th and 8th graders in Math and Reading in 2005. Again using the example in Figure 1, State A has a strictness of negative ten while State B's strictness is positive ten.

Given the different grades and subjects, the sample size was almost 4 times the number of states with available data (N=184). With few exceptions, more students passed the state test than the NAEP test. As in the example above, the more negative the difference between the percent proficient on the NAEP and state exams, the easier the test relative to the student level in that state. Where states did not test 4th and 8th graders in 2005, (states originally had to test at least two grades between 3 and 8), I imputed the percent of students that passed the grade 4 exam from grades 3 and 5 and grade 8 from 7. To impute these percentages I took states that tested grades 3, 4 and 5 and regressed the percentage of students that passed in grade 4 for that subject on

grades 3 and 5. Similarly, I regressed the percentage of students that passed in grade 8 on the percentage that passed in grade 7 in states that tested both grades to find their correlation. With these coefficients, I predicted scores for grade 4 in states with grade 3 and/or 5 and grade 8 with grade 7 in those states that only tested grade 7.

To answer the second question concerning whether NAEP and state scores moved together over time, I calculated the correlation between the percent of students proficient on state and NAEP tests from 2005 to 2007.

To test the hypothesis that strictness of state tests was associated with larger NAEP gains (question 3), I regressed the change in NAEP scores from 2003 to 2007 on state strictness in 2005:

$$Y_{2007N} - Y_{2003N} = \beta_0 + \beta_1(P_{2005N} - P_{2005S}) + \epsilon$$

Y_{2007N} = percent of students proficient on NAEP in 2007

Y_{2003N} = percent of students proficient on NAEP exam in 2003

P_{2005N} = percent of students proficient on NAEP in 2005

P_{2005S} = percent of students proficient on State exam in 2005

I added a dummy for grade and subject to this regression:

$$Y_{2007N} - Y_{2003N} = \beta_0 + \beta_1(P_{2005N} - P_{2005S}) + \beta_2G + \beta_3S + \epsilon$$

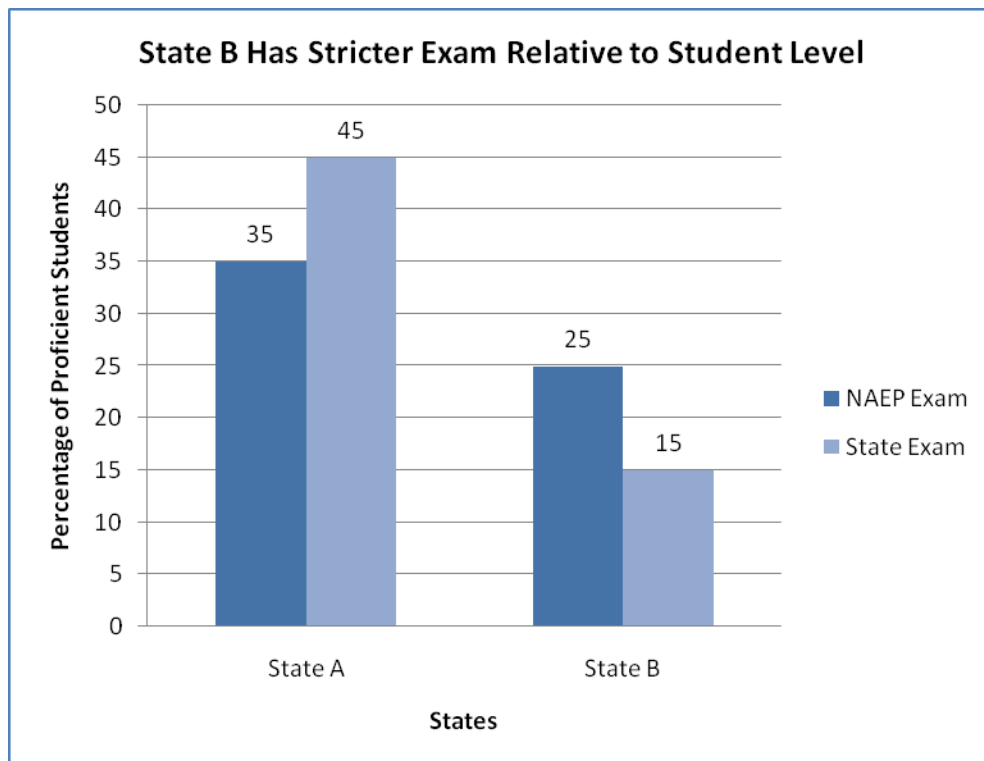
G = Grade8

S = Math

I estimate a regression of absolute NAEP proficiency in 2003 and 2007 (rather than change over time) on state strictness in 2005 including a dummy variable for year. This regression examines the level of proficiency and strictness rather than the relationship between progress and strictness. The interpretation of this coefficient is quite similar to that of the previous model.

For each of these regressions, I ran a fixed-effects regression to see if this effect held when using only variation within states. I ran these regressions without the missing data for grades 4 and 8 and then

Figure 1



with the imputed percentages using a linear regression model in all cases.

RESULTS

Question 1: *Did standards differ among states in 2005?*

As expected, there was great variation in state standards in 2005. The most challenging state exam relative to the NAEP was the 8th grade exam in Missouri, the Missouri Assessment Program (MAP, not to be confused with the national MAP exam). In 2005, 26 percent of 8th graders passed the NAEP while only 16 percent were able to pass MAP, a positive 10 percent difference (see Figure 2). The least difficult state exam relative to student levels was the 4th grade reading exam in Mississippi. In 2005, 89 percent of Mississippi students passed their state exam while only 18 percent were able to pass the NAEP exam, a difference of 71 percentage points.

Across all states on average, 35 percent more students passed the state exams than passed the

NAEP exam in that grade and subject. Over both grades and subjects in 2005, Massachusetts had the hardest relative tests, with 2 percent more students passing NAEP than the state test; Tennessee had the easiest relative state tests with 61 percent more students passing state tests than the NAEP.

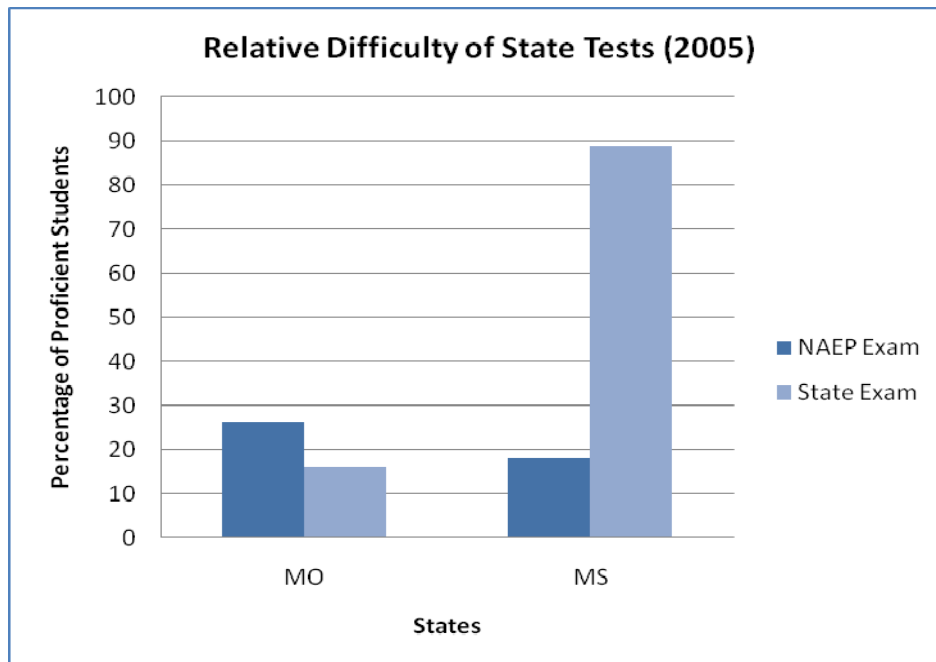
Question 2: *Did state standards change from 2005 to 2007?*

Table 1

Correlations between State Test and NAEP Scores		
	Reading	Math
Grade 4	-0.07	0.05
Grade 8	-0.24	0.19

Not only were the relative state standards different in 2005 but these standards changed from 2005 to 2007. Generally, states showed much faster improvement on state exams than NAEP exams. The percentage of students passing state tests and NAEP did not show strong correlation from 2005 to 2007 and was in fact negative for reading in 4th and 8th grade. None of these correlations were statistically

Figure 2: A Difference in Standards



different from 0, even at the 10% level.

Greater improvement on state tests did not mean greater improvement on the NAEP than other states, but in the case of reading exams, the point estimates suggest less improvement on NAEP proficiency.

That is, states that showed above-average increases in the percent of students passing their state reading test across states would more often have below-average improvement on NAEP. Figure 3 illustrates state trends in improvement from 2003 to 2007 on 4th grade state and NAEP reading tests.

In Massachusetts, the percent of students passing the 4th grade NAEP reading assessment increased from 2003 to 2007 by almost 10 percentage points.

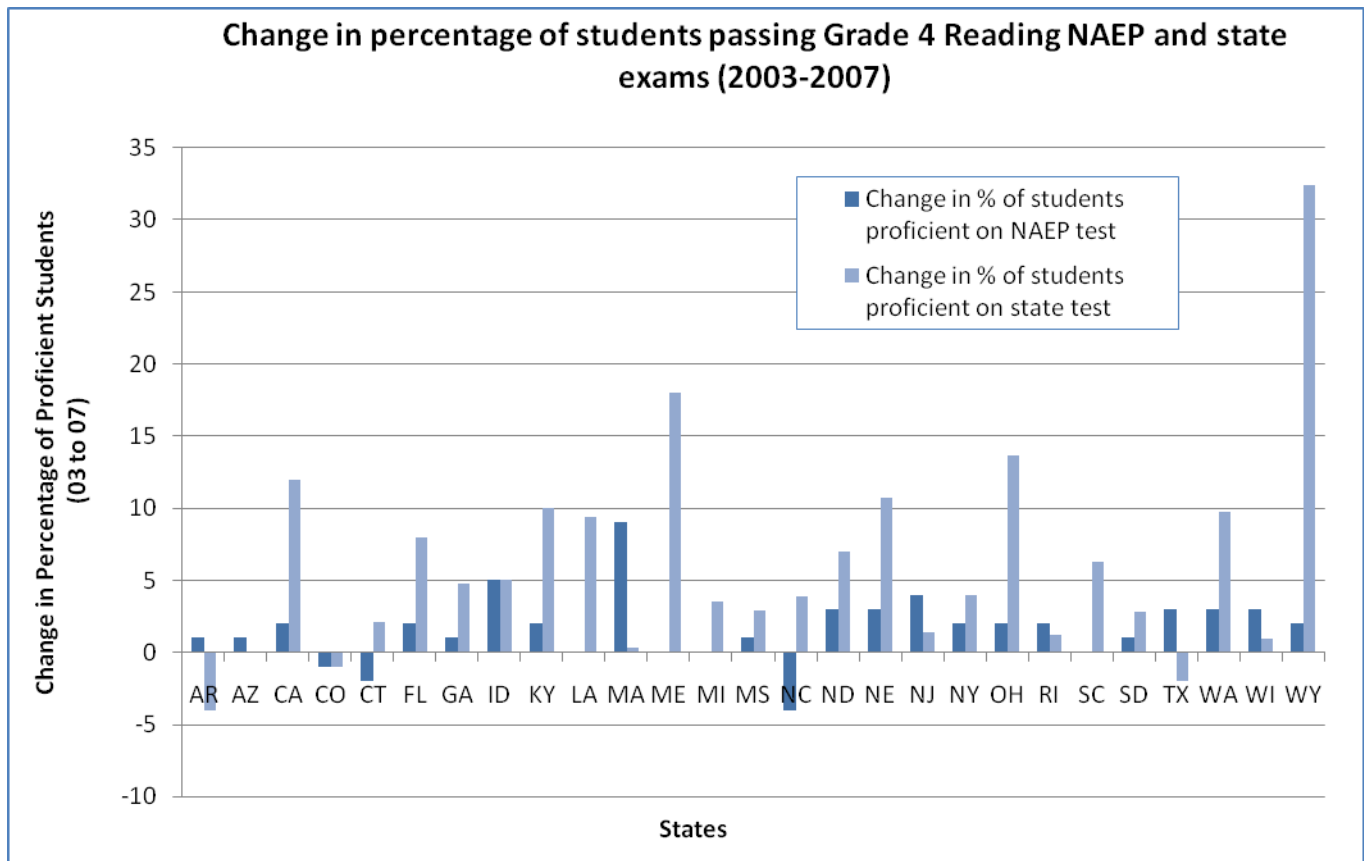
During this period, there was almost no increase in the percent of students scoring proficient on the Massachusetts state exam for grade 4 reading. I

argue that Massachusetts increased the rigor of its standards as opposed to states like Wyoming and Maine that lowered their standards. In the case of Maine there was no increase in the percent of students that passed the NAEP. Wyoming had only a two-percentage point increase on the NAEP but quite a large increase in the percent of students who passed the state exams. This is an illusory gain.

Question 3: Did higher state standards in 2005 correlate with greater improvement on a national exam?

More importantly, states that maintained high standards in 2005 generally saw greater improvement on average in NAEP proficiency from 2003 to 2007. A simple regression of change in NAEP proficiency on state strictness (Figure 4 and

Figure 3: Gains: Real or Illusory?



* See Appendix Figures 1A-C for summary of Grade 4 Math, and Grade 8 Reading and Math

Figure 4: A Strong Positive Correlation

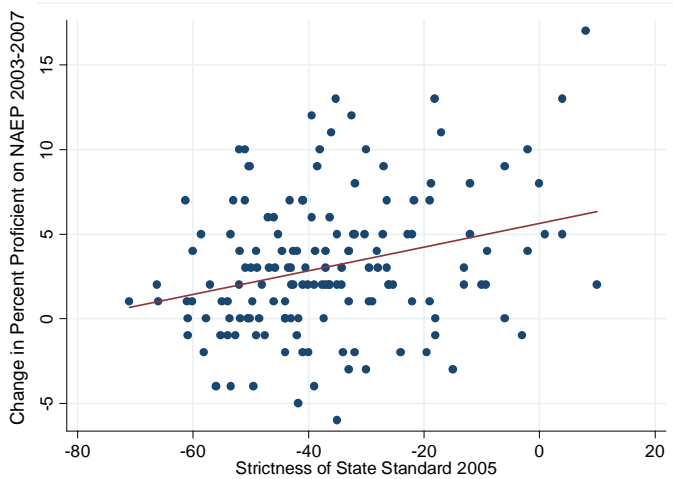


Table 2) shows that given an increase in relative strictness of one percent, we would expect 0.07 percent more students in the state to pass the NAEP. The coefficient is statistically different from zero at the one percent level suggesting that the observed correlation is not due to sampling error.

The size of the coefficient implies that a state with a test that gave pass rates 10 percent above the NAEP that increased the stringency of its test so that its test gave pass rates 10 percent below the NAEP could see a 1.4 percentage point gain in pass rates on the NAEP over four years as a result. This may seem like a small effect, but when compared with costly interventions that produce similar size effects, the effect seems quite large.

Inclusion of subject and grade indicators in the regression shows some interesting patterns (Table 3-Model 2). The math dummy variable was highly significant and positive while the grade 8 dummy variable was highly significant and negative. Given the same state strictness and grade level, on average the change in the percentage of students passing the NAEP math exam from 2003 to 2007 increased by 4.7 percent more than on reading. The negative coefficient for grade 8 means that on average, the change in the percentage of students passing NAEP

from 2003 to 2007 at the 8th grade level was 3.6 percent lower than at the 4th grade level.*

Regressing absolute NAEP proficiency on state test strictness in 2005, and dummy variables for year, grade and subject, yielded highly significant results for strictness with a coefficient of .09 (Table 4). Recall that this regression does not measure improvement on NAEP, but rather the absolute level of NAEP proficiency before and after 2005. The

strength of the strictness coefficient indicates that states with higher NAEP proficiency before and after 2005 were also more likely to have a stricter exam in 2005. The coefficient for the year variable was highly significant and strong at .77 showing that NAEP scores were higher in 2007 than 2003 controlling for strictness, grade, and subject.

Table 2

Change in NAEP Proficiency 2003-2007		
Variable	Coef.	P>t
Intercept	5.64	0 **
State strictness	0.07	0 **
Prob > F	0 **	
R-squared	0.08	
Adj R-squared	0.07	
Observations	153	**p<=.01

Despite a lower coefficient on strictness, the fixed-effects model was significant at the 10% level for NAEP proficiency in 2003 and 2007 indicating that this effect holds when based on within state variation (Table 5). However, when regressing the *change* in NAEP scores on strictness in the fixed-effects

* The significance of the strictness coefficient and the dummy variables holds with imputed values. See Table 1 of Appendix for results that include imputed values.

Table 3

Change in NAEP Proficiency 2003-2007				
Variable	Model 1 [^]		Model 2 ^{^^}	
	Coef.	P>t	Coef.	P>t
Intercept	9.22	0.00 **	4.37	0.00 **
State strictness	0.09	0.00 **	0.05	0.00 **
Grade 8	-3.75	0.00 **	-3.65	0.00 **
Math			4.70	0.00 **
Prob > F	0 **		0 **	
R-squared	0.28		0.59	
Adj R-squared	0.27		0.58	
Observations	153		153	

[^]includes dummy variable for grade
^{^^}includes dummy variables for grade and subject **p<=.01

model, strictness was not significant. This suggests that there is variation in the effect of strictness on improvement in NAEP proficiency across states. This is likely the result of different policies across states.

Table 4

Percent Proficient on NAEP in 2003 & 2007		
Variable	Coef.	P>t
Intercept	-1510.19	0.00 **
State strictness	0.09	0.00 **
Grade 8	-3.97	0.00 **
Math	0.33	0.69
Year	0.77	0.00 **
Prob > F	0.00	
R-squared	0.58	
Adj R-squared	0.59	
Observations	306	**p<=.01

CONCLUSIONS

We do not know why states with harder tests did better in terms of making real gains on the NAEP exam, nor do we know why states like Massachusetts raised the standards of their state

exam. Higher standards may act as an incentive to improve performance (by focusing public attention on proficiency) or the information that states gain from rigorous standards may provide a better framework for efficiently allocating resources. The informational advantage theory would likely predict a slower impact of standards on real gains than the incentive theory, but in practice both theories are likely to explain some performance gains and the two explanations may interact in unpredictable ways. States that have harder tests may be doing other things that improve performance. We can only identify states that perform well while maintaining high standards and states that take advantage of flexible NCLB guidelines.

Recognizing the differences in the difficulty of state exams understates the extent to which states are positioning themselves to escape accountability under NCLB. Not only did states have different standards in the years for which I collected data, 2003, 2005 and 2007, but test standards were apparently adjusted over time so that some states did not improve at all on the NAEP but showed vast improvement on their state test.

Table 5

Fixed-Effects Models				
Change in Percent Proficient on NAEP 2003-2007			Percent Proficient on NAEP in 2003 & 2007	
Variable	Coef.	P>t	Coef.	P>t
Intercept	2.36	0.08 *	-1511.36	0.00 **
State strictness	0.00	0.98	0.06	0.06 *
Grade 8	-3.48	0.00 **	-3.84	0.00 **
Math	5.08	0.00 **	0.52	0.23
Year	(dropped)		0.77	0.00 **
State	F(43, 106)	0.00 **	F(43, 258)	0.00 **
Prob > F	0.00	**	0.00	**
R-squared	0.81		0.85	
Adj R-squared	0.73		0.82	
Observations	153		306	*p<= .1 **p<=.01

Without accurate information, parents may not know whether their child’s school is passing standards that are considered basic in another state, and states cannot target districts that need greater resources. The analysis, however, is promising in that those states that set higher standards performed slightly better over time on a standardized national exam. This effect seemed even stronger on math exams and at the fourth grade level. Though we cannot infer any causal relation between standards and improvement, their correlation suggests that challenging standards or the accountability that accompanies high standards may raise student achievement.

In its reauthorization of No Child Left Behind, Congress should consider the merits of a national exam somewhat like the NAEP or creating incentives to align state tests with NAEP standards. Though federal intervention involving a single exam might face legal obstacles, developing a national exam that is both realistic and challenging would

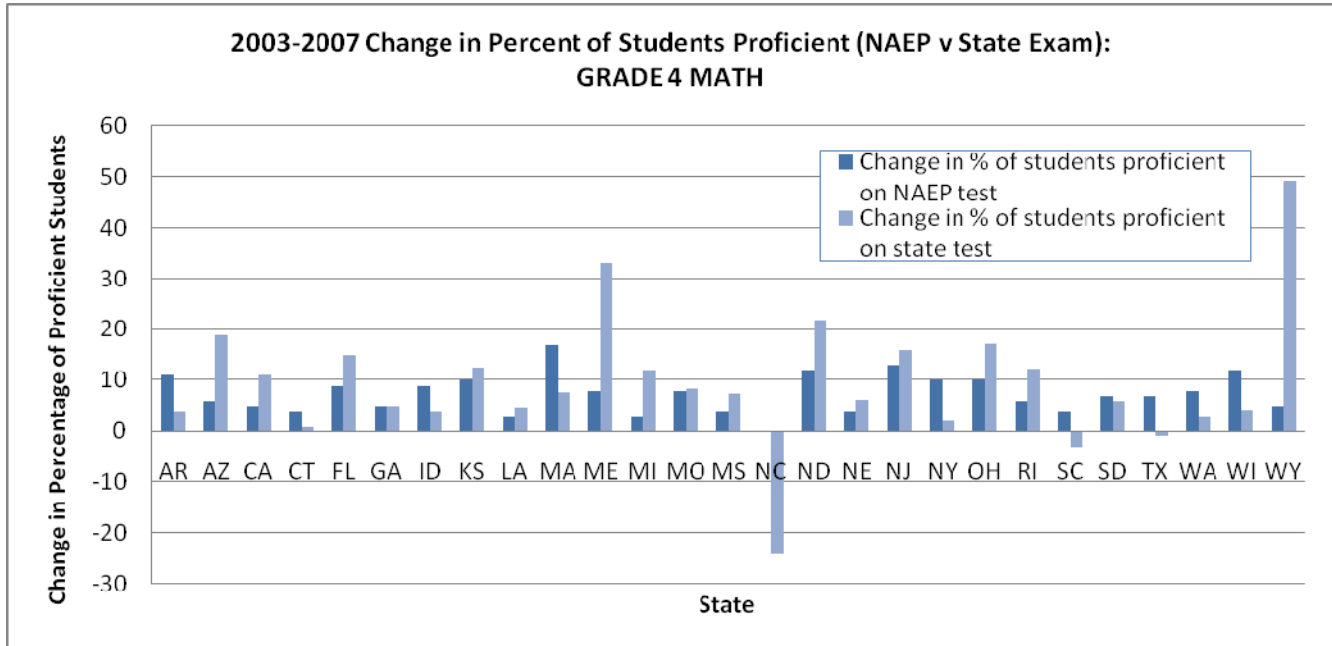
ensure that consistent standards exist for evaluating improvements in the American school system.

To avoid punishing school districts with limited resources, NCLB must continue to provide attention and resources to underperforming schools through the expansion of Title 1 funds. Successful strategies should be applied to states and school districts with similar socioeconomic and demographic characteristics.

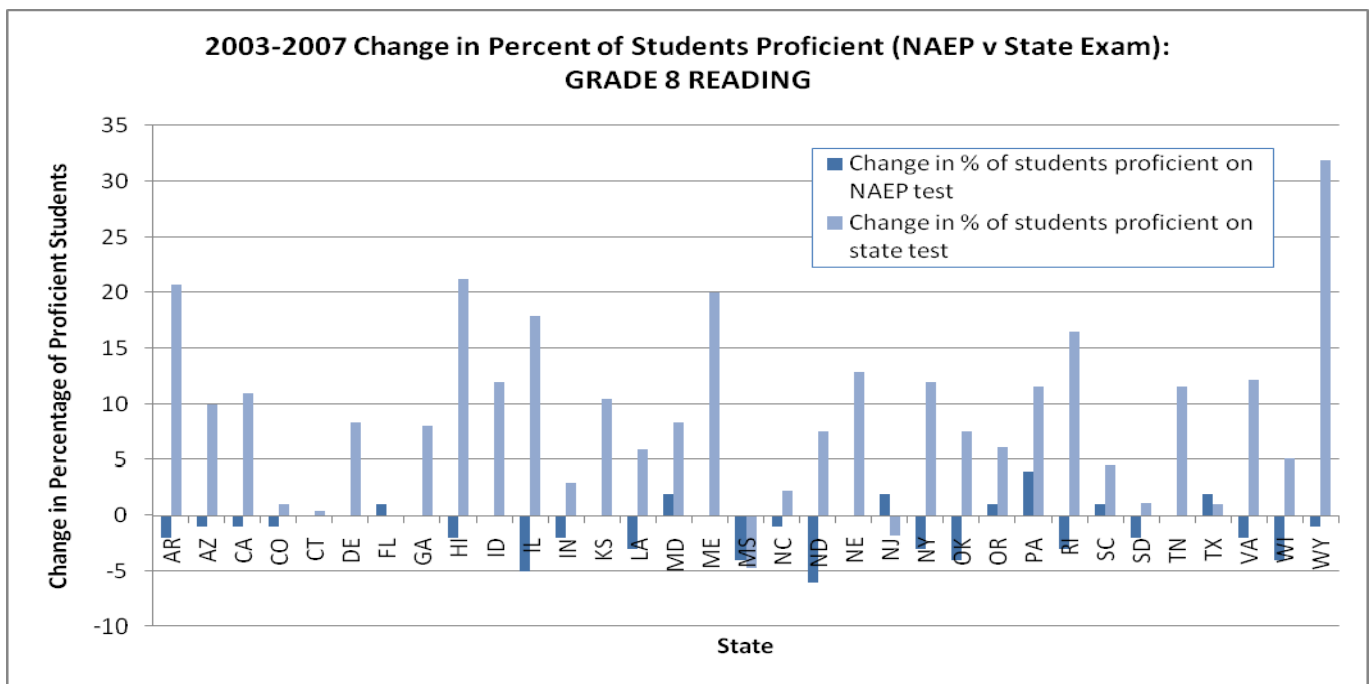
Further work should be done studying the characteristics of those states that lowered their standards the most. Are they at a financial disadvantage or were their state tests too difficult in 2003 for states to meet the goal of 100 percent proficiency by 2014 without altering the test? These questions could be analyzed quantitatively with the inclusion of variables for state median income and state budget. Also warranted is further investigation into why high standards in 8th grade were not as effective as in 4th grade and why students seemed to improve more in math than in reading.

APPENDIX TABLES

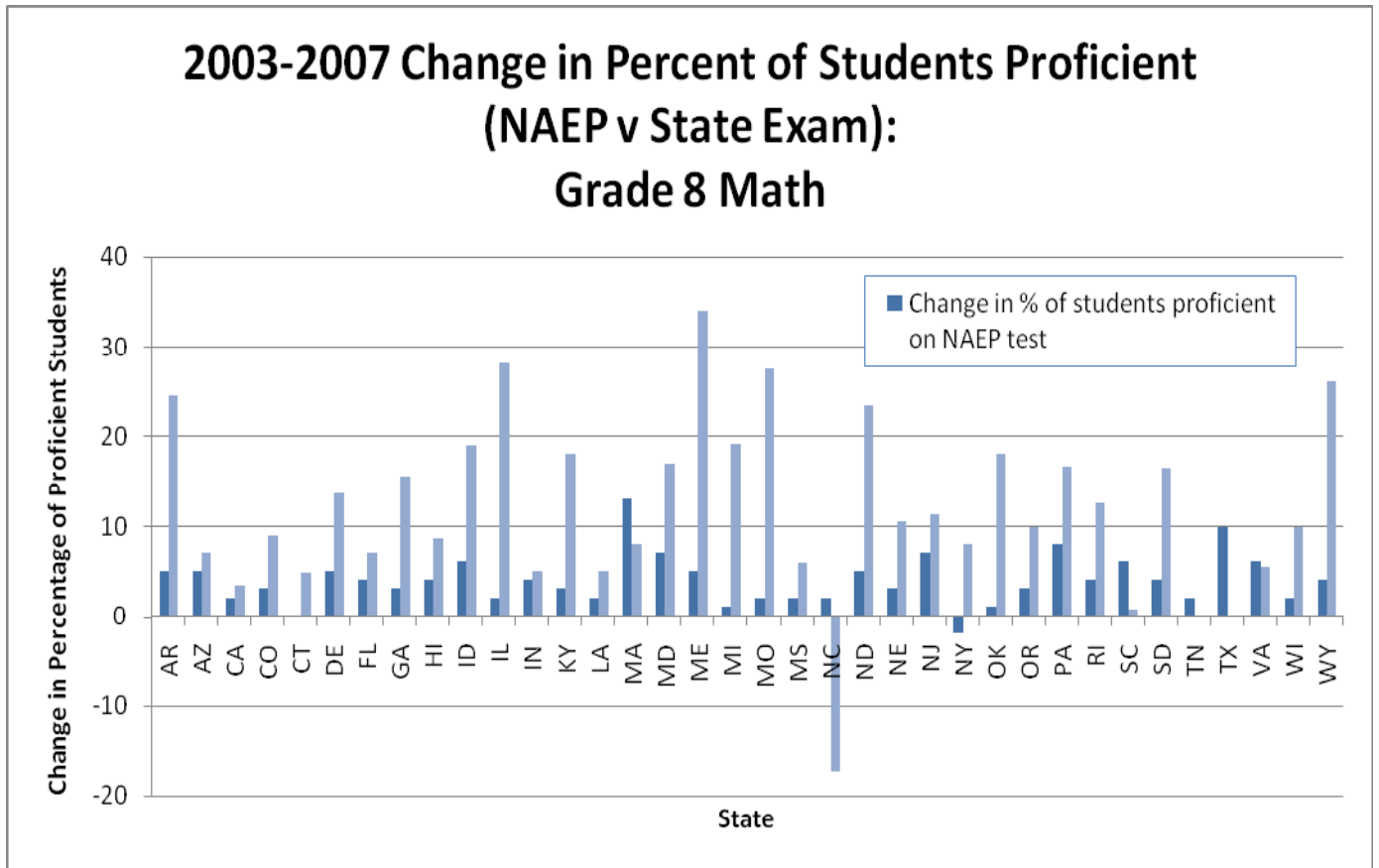
Appendix Figure 1-A



Appendix Figure 1-B



Appendix Figure 1-C



Appendix Table 1

Models Including Imputed Data						
Change in Percent Proficient on NAEP 2003-2007				Percent Proficient on NAEP in 2003 & 2007		
Variable	Coef.	P>t		Coef.	P>t	
Intercept	4.38	0.00	**	-1492.62	0.00	**
State strictness	0.05	0.00	**	0.09	0.00	**
Grade 8	-3.76	0.00	**	-3.78	0.00	**
Math	4.77	0.00	**	0.61	0.42	
Year	(dropped)			0.76	0.00	**
Observations	178			356		
Prob > F	0.00	**		0.00	**	
R-squared	0.58			0.13		
Adj R-squared	0.59			0.12		**p<=.01

REFERENCES

- ¹ Mullis, Ina V.S., Michael O. Martin, E.J. Gonzalez, and S.J. Chrostowski. "Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades." Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College, 2004.
- ² Mullis, Ina V.S., Michael O. Martin, Ann M. Kennedy, and Pierre Foy. "PIRLS 2006 International Report." Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2007.
- ³ Organization for Economic Cooperation and Development. "PISA 2006: Science Competencies for Tomorrow's World Executive Summary," 2007.
- ⁴ 2000 US Census Bureau
- ⁵ LeFloch, Kerstin Carlson, Felipe Martinez, Jennifer O'Day, Brian Stecher, James Taylor, Andrea Cook. "State and Local Implementation of the No Child Left Behind Act Vol. 3. Accountability Under NCLB-Interim Report." 2. US Department of Education, 2007.
<http://www.air.org/publications/documents/NCLBAccountabilityReport.pdf>
- ⁶ Rueben, Kim, Sheila E. Murray. "State Education Spending: Current Pressures and Future Trends." 327. National Tax Journal, 2007.
- ⁷ California Department of Education. "Charter Schools FAQ Section 12: Questions relating to No Child Left Behind issues." www.cde.ca.gov/sp/cs/re/qandasec12dec05.asp
- ⁸ California Department of Education. *ibid.*
- ⁹ King, Kent D. "Missouri Consolidated State Application Accountability Workbook for State Grants under Title IX, Part C, Section 9302 of the Elementary and Secondary Education Act (Public Law 107-110)." 16. 2003/2008. www.ed.gov/admins/lead/account/stateplans03/mocsa.pdf
- ¹⁰ Scott, Caitlin, Maureen Kelleher. "The Sit-Down Dinner. Formalizing Restructuring Under the No Child Left Behind Act in Michigan." 6. Center on Education Policy. 2007. www.cep-dc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documentid=237&documentFormatId=3750
- ¹¹ LeFloch, Kerstin Carlson. *ibid.* 5.
- ¹² Scott, Caitlin. *ibid.* 9.
- ¹³ Carey, Kevin. "Hot Air: How States Inflate Their Educational Progress Under NCLB." Education Sector. 5. 2006. www.educationsector.org/usr_doc/Hot_Air_NCLB.pdf
- ¹⁴ Chudowski, Naomi and Vic Chudowski. "Many States Have Taken a "Backloaded" Approach to No Child Left Behind Goal of All Students Scoring 'Proficient'." 1. Center on Education Policy. 2008. www.cep-dc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documentid=238&documentFormatId=3756
- ¹⁵ Arizona Education Association. "Politics and Legislation: NCLB Historical Timeline." 2008. http://www.arizonaaea.org/politics.php?page=282&sub_page=252

- ¹⁶ Cronin, Joe, Michael Dahlin, Deborah Adkins, G. Gage Kingsbury. “The Proficiency Illusion.” 36. Thomas B Fordham Institute. 2007. www.edexcellence.net/doc/The_Proficiency_Illusion.pdf
- ¹⁷ Cronin, Joe. *ibid.* 16.
- ¹⁸ Vu, Pauline. “Do state tests make the grade?” Stateline. 2008. <http://www.stateline.org/live/details/story?contentId=272382>
- ¹⁹ Cronin, Joe. *ibid.* 13.
- ²⁰ Carey, Kevin. *ibid.* 5.
- ²¹ Carey, Kevin. *ibid.* 6.
- ²² King, Kent D. 18, 19.
- ²³ Carey, Kevin. *ibid.* 6.
- ²⁴ US Department of Education. Federal Register: Rules and Regulations. Vol. 72 No. 67. 2007. <http://ed.gov/legislation/FedRegister/finrule/2007-2/040907a.html>
- ²⁵ Carey, Kevin. *ibid.* 8.
- ²⁶ Carey, Kevin. *ibid.* 10.
- ²⁷ LeFloch, Kerstin Carlson. *ibid.* 21.
- ²⁸ Carey, Kevin. *ibid.* 7.
- ²⁹ Kober, Nancy, Naomi Chudowsky, Victor Chudowsky. “Has Student Achievement Increased since 2002? State Test Score Trends Through 2006–07.” 2008. <http://www.cep-dc.org/index.cfm?fuseaction=Page.viewPage&pageId=495&parentID=481>



THOMAS GONZALEZ

Is currently a senior at the University of Chicago majoring in Economics. Mr. Gonzalez is a member of the 2008 class of the UI Summer Academy.

RESEARCH MENTORS

AUSTIN NICHOLS

Dr. Nichols is a Senior Research Associate in the Income and Benefits Policy Center at the Urban Institute.

DOUG WISSOKER

Dr. Wissoker is a Senior Research Associate in the Urban Institute’s Statistical Methods Group.

COLIN TAYLOR

Mr. Taylor is currently a graduate student at Harvard University. He was formerly a Research Assistant in the Urban Institute’s Education Policy Center.

The Summer Academy is an initiative of the **Urban Institute** created to address the underrepresentation of minorities and persons from distressed communities in public policy research. It is an intensive eight-week program of technical skills training and career exposure targeted to promising college juniors. The program takes place on location at the Urban Institute in Washington, D.C. For more information, go to <http://summeracademy.urban.org>.



The Urban Institute

GATHERS data,
CONDUCTS research,
EVALUATES programs,
OFFERS technical ASSISTANCE overseas, and
EDUCATES Americans on social and economic issues — to foster sound public policy and effective government.

2100 M Street NW
 Washington, DC 20037
www.urban.org