REL APPALACHIA
Regional Educational Laboratory
At CNA Corporation

Assessing the likelihood that Virginia schools will meet the proficiency goals of the No Child Left Behind Act

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

ISSUES & ANSWERS

REL 2007–No. 012

REL APPALACHIA
Regional Educational Laboratory
At CNA Corporation

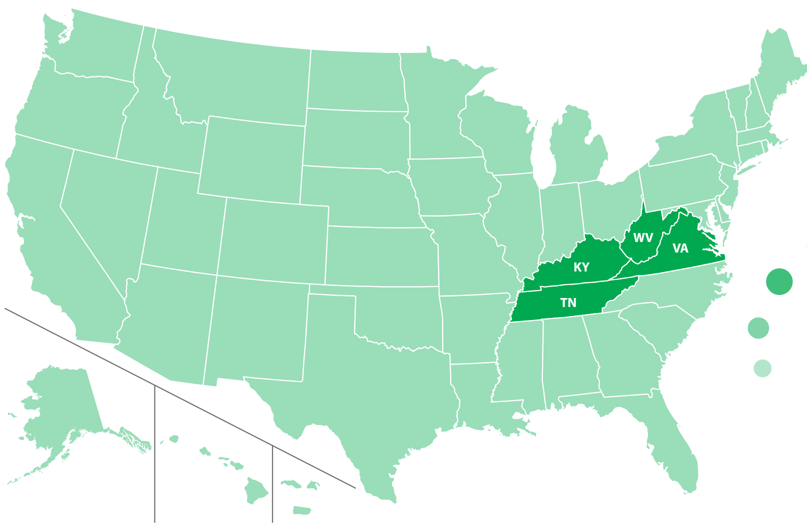
Assessing the likelihood that Virginia schools will meet the proficiency goals of the No Child Left Behind Act

September 2007

Prepared by

Louis Jacobson, Ph.D.
The CNA Corporation

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education



REL
APPALACHIA
At CNA Corporation

Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

September 2007

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0021 by Regional Educational Laboratory Appalachia administered by The CNA Corporation. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Jacobson, L. (2007). *Assessing the likelihood that Virginia schools will meet the proficiency goals of the No Child Left Behind Act* (Issues & Answers Report, REL 2007–No. 012). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from http://ies.ed.gov/ncee/edlabs

This report is available on the regional educational laboratory web site at http://ies.ed.gov/ncee/edlabs.

Summary

Assessing the likelihood that Virginia schools will meet the proficiency goals of the No Child Left Behind Act

This report investigates progress in Virginia public schools in satisfying the requirement of the No Child Left Behind Act of 2001 that every student be proficient in reading and math by 2014. It develops a variable change model that uses observed baseline proficiency and proficiency trends at individual schools to forecast gains for six subgroups in elementary, middle, and high schools.

The study finds that there were substantial increases in proficiency overall and especially large increases for schools and subgroups that had low proficiency levels in 2002. The forecasts indicate that there will continue to be substantial proficiency increases in the near term, but that with few exceptions proficiency will plateau at levels well below 100 percent before 2014.

The report looks at proficiency levels for students overall and in six subgroups in elementary, middle, and high schools in reading and math in more than 1,600 Virginia schools in the first four years after passage of the No Child Left Behind Act. These observed trends are then used to forecast reading and math proficiency levels for 2006 through 2014.

The forecast of proficiency levels relies on a variable change model rather than a constant change model. The model is based on the assumption that as a school reaches higher proficiency levels, its proficiency will grow at the average annual 2002–05 pace achieved by other schools of the same type that attained that higher base proficiency level in 2002. Estimating the relationship between the change in proficiency between 2002 and 2005 and the proficiency level attained in 2002 provides a realistic answer to the central analytic question: Are rates of improvement likely to rise, fall, or remain constant relative to current rates?

The research yielded several major findings:

- Virginia schools increased their proficiency levels between 2002 and 2005. In 2002, in an average school, 74.3 percent of students tested proficient in reading. On average, a school's reading proficiency increased by 6.9 percentage points over the next three years (7.0 for elementary schools, 7.6 for middle schools, and 5.9 for high schools). In an average school math proficiency increased by 9.8 percentage points (9.2 for elementary schools, 9.0 for middle schools, and 13.0 for high schools).

- There was substantial variation in reading proficiency change across schools of each type between 2002 and 2005. Seventy percent of elementary schools had changes

of between –3.3 and +17.3 percentage points, 70 percent of middle schools had changes of between –1.2 and +16.4 percentage points, and 70 percent of high schools had changes of between –3.1 and +14.9 percentage points. Variation in math proficiency was also large. Approximately 70 percent of elementary, middle, and high schools had changes of between –2.0 and +20.4 percentage points.

- Improvements in proficiency are likely to continue but at a reduced pace. For example, in elementary schools African American reading proficiency averaged between 40 and 50 percent in 2002, and proficiency increased by about 20 percentage points over the next three years. In an average school African American proficiency is forecast to increase by about 10 percentage points between 2005 and 2007 and by about 5 percentage points between 2008 and 2010.

- Actual and forecast increases in proficiency were greatest for schools and subgroups that attained low levels in 2002. For example, in an average high school reading proficiency for students with disabilities was 46 percent in 2002 and increased by 14 percentage points by 2005. In contrast, in an average high school reading proficiency levels for students without disabilities averaged 86 percent in 2002 but increased by only 5 percentage points by 2005.

- It becomes increasingly difficult to boost proficiency as proficiency levels rise. For example, subgroups in schools with 80 percent or higher average proficiency levels in 2002 exhibited declines over the next three years. Annual changes in proficiency declined by about 2.2 percentage points for every 10 percentage point increase in proficiency, and this trend is likely to continue into the future.

These findings lead to three major conclusions:

- The actual and forecast performance of Virginia public schools is in keeping with the intent of the No Child Left Behind Act to substantially increase proficiency levels, especially for schools and subgroups with low initial levels.

- Despite the strong observed and forecasted gains, it is unlikely that most schools will show consistent gains beyond 95 percent proficiency levels for whites; beyond 80 percent for African Americans, Hispanics, students with limited English proficiency, and economically disadvantaged students; and beyond 70 percent for students with disabilities.

- Close to 100 percent of Virginia's schools will not meet the status standard in 2014 as it rises from 69 percent in 2006 to 100 percent in 2014. However, it is difficult to determine how not meeting the status standard translates into schools being labeled in need of improvement. This is because the alternative "safe harbor" standard (which in Virginia is based on the year to year increase in the percentage of students testing proficient) could save as many as half of the schools from becoming identified as needing improvement.

The research also reached four methodological conclusions related to forecasting proficiency:

- The models typically used to forecast changes in proficiency are based on the untenable assumption that a school's observed increase over a base period will continue unchanged into the future. These models also do not accurately describe the range of outcomes across schools, but instead focus on the performance of schools with average characteristics.

- In forecasting proficiency it is essential to model the slowdown in a school's proficiency change as the school's proficiency level rises, to describe the range of changes across schools with similar proficiency levels, and to take into account the statistical error of the estimates.

- A school's year-to-year fluctuations in proficiency level can have a major effect on meeting performance standards, because those fluctuations are frequently large enough to allow the school to meet safe harbor standards. Modeling the meeting of safe harbor standards was outside the scope of this study but is essential for predicting which schools will be labeled as needing improvement.

- Year-to-year fluctuations in proficiency levels of individual Virginia schools were so large that it was difficult to discern long-term trends in data covering four years. This finding has important implications for developing performance measures and standards that accurately reflect progress.

September 2007

TABLE OF CONTENTS

Tables

This report investigates progress in Virginia public schools in satisfying the requirement of the No Child Left Behind Act of 2001 that every student be proficient in reading and math by 2014. It develops a variable change model that uses observed baseline proficiency and proficiency trends at individual schools to forecast gains for six subgroups in elementary, middle, and high schools.

OVERVIEW

This report provides information about progress in meeting the central proficiency goal of the No Child Left Behind Act and develops a method to accurately forecast proficiency change. It lays the basis for future work on accountability systems for the No Child Left Behind Act, a highly complex topic (see box 1 for details of the No Child Left Behind Act and its application in Virginia).

The analysis stems from a request by Virginia education officials to develop a means of forecasting changes in school proficiency that would enable the state to better predict which schools are likely to have problems meeting accountability standards and, of these, which are having problems that are unlikely to be adequately resolved without further action. The ultimate goal is to create an early warning system for the state, districts, and schools.

A key element of the work was developing an accurate way to deal with a statistically complex issue—modeling how a school's proficiency level will change over time. The complexity stems from the large variation across schools at any one time and the variation across individual schools over time.

The literature on forecasting change in school-level proficiency does not capture the complexity observed in Virginia. Thus, much of the work for this report was aimed at improving forecasting methods by taking the variation into account and by capturing the relationship between a school's proficiency level and the way that level changes as higher proficiency levels are reached.

The report presents details of the analysis so that technical experts can review the basis of the conclusions and so that other researchers can replicate the analysis using data from other states.

A comprehensive database from the Virginia Department of Education (see box 2) enabled the investigation of two key research questions:

BOX 1

The No Child Left Behind Act of 2001 and its application in Virginia

The No Child Left Behind Act of 2001 became law in January 2002. Its purpose is to ensure that "all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments" (section 1001).

The law requires states to develop a testing program to determine the percentage of students in a school who are proficient in reading and math. States have wide discretion in specifying the tests and the scores required to be considered proficient. States and districts that do not develop accountability systems and do not impose sanctions on persistently underperforming schools can lose federal funds.

Virginia's testing system

Virginia was one of 17 states that had developed an accountability system in response to the 1994 Elementary and Secondary Education Act. Its comprehensive Standards of Learning testing program was first used in classrooms in 1998 (Virginia Department of Education, 2000). To meet the requirements of the No Child Left Behind Act, Virginia used its existing Standards of Learning to assess reading and math proficiency. Virginia's testing program is among the highest rated in the nation (Council of Chief State School Officers, 2006.) In particular, Virginia gets high marks for the rigor of the tests and the high proficiency standards to which students are held.

From 2002 through 2005 Virginia's No Child Left Behind testing program included math and reading/ English language arts tests given in grades 3, 5, and 8; end-of-course tests for Algebra I, Algebra II, and Geometry for middle and high school students; and a comprehensive high school English test, usually given in grade 11.[1] Students who did not test as proficient on end-of-course tests could retake the tests. If they subsequently tested as proficient, that success was counted, but the failures remained on the schools' records. Test results were excluded for limited English proficiency students in their first year in a U.S. school as well as for some transfer students.

Adequate yearly progress and safe harbor

States have to determine annually whether a school is making adequate yearly progress, mainly by establishing whether a school meets or exceeds the state-set status (level) standard for the percentage of students testing as proficient in reading and math in the aggregate and for up to six subgroups. This standard is called an annual measurable objective. In addition, to make adequate yearly progress, 95 percent of students must be tested in grades where testing occurs. The annual measurable objective must reach 100 percent by 2014. In Virginia the annual measurable objective was set at 60.7 percent in 2002 and rose to 65 percent by 2005. It will rise by 4 percentage points a year until 2013 and then by 3 percentage points in 2014 to 100 percent.

To determine whether schools are making adequate yearly progress, Virginia requires measuring proficiency for up to six subgroups—whites, African Americans, Hispanics, students with limited English proficiency, students from economically disadvantaged families, and students with disabilities. For a school to be held accountable for a subgroup a minimum of 50 tests must have been taken by members of the subgroup in a given year.

No Child Left Behind allows states to apply an alternative "safe harbor" (growth) standard when schools do not make adequate yearly progress based on the status (level) standard. Under safe harbor, schools are considered to make adequate yearly progress even if they do not meet annual measurable objectives if they reduce by at least 10 percent the proportion of students who scored below proficient the previous year.

Because of the safe harbor provisions, even if a school falls far below the status standard, it might not consistently fail to make adequate yearly progress. Moreover, under the No Child Left Behind Act schools must fail to make adequate yearly progress for two successive years before being labeled "in need of improvement," which is when sanctions begin to be applied for each additional consecutive year of failure. The sanctions progress from requiring improved planning and offering students the chance to transfer to other schools to wholesale changes in staffing and governance. Ultimately, teachers, principals, and superintendents can lose their jobs. This feature of the No Child Left Behind Act creates a high-stakes testing system.[2]

Notes

1. In accordance with No Child Left Behind requirements Virginia began testing students in grades 4, 6, and 7 in 2006.
2. Originally, this study intended to forecast rates for schools being labeled in need of improvement as well as for those failing to meet proficiency status standards, but forecasting failure rates proved to be exceptionally complex because of safe harbor provisions and the requirement to not make adequate yearly progress in two successive years before being labeled in need of improvement.

BOX 2

The database

A large and detailed database provided by the Virginia Department of Education was used to examine proficiency in Virginia schools during the first four academic years after passage of the No Child Left Behind Act of 2001. The database lists each school's proficiency level in math and reading and the number of math and reading tests taken in each school for each year 2002 through 2005. Each school's data are also disaggregated into the six subgroups for which Virginia's schools are held accountable.

While the database covers all 1,842 Virginia public schools operating at any time from 2002 through 2005, the results for all students together are based on data from 1,601 schools, which reported reading results for 194 students on average. About 50 public schools were dropped because they were run at the state or regional level, including regional vocational schools, specialized governor's schools for talented students, and schools for prisoners or for students in state-run medical facilities. The analysis also omitted data for 121 schools that were not in operation during each year 2002 through 2005 (usually because they opened during this period) and 48 elementary schools that did not include both the third and fifth grades—the elementary grades in which No Child Left Behind testing occurred in Virginia. Also omitted were 72 "mixed" schools containing grades spanning two or more types of schools.[1] These omissions ensured that the performance of a single cohort of schools with comparable data for each year was being examined. The accuracy of the forecasts would have suffered if schools entered or left the sample and did not have proficiency measures for at least four years.

Detailed results for reading, disaggregated by subgroup, are included in the main report because the data clearly showed that proficiency increased at a slower rate for reading than in math among schools at the same proficiency level in 2002. Thus, attaining the key goal of 100 percent proficiency hinges on overcoming reading proficiency shortfalls. Only the plateau point forecasts for math are covered in the main report.

Note

1. Initially, results were broken down for elementary, middle, high, and "other" schools. However, the "other" group was dropped because its omission did not materially change the results for all schools together, and in the disaggregated tables the results for the "other" group were nearly identical to the results for the three main school types weighted to reflect the mix of grades included in the "other" schools.

1. What are the distribution across Virginia schools of proficiency levels in 2002 and the changes in proficiency attained by 2005 for all students together and for each of six subgroups?

2. How will the distribution of proficiency levels change from 2006 through 2014?

The 2006–14 forecasts of school proficiency levels rely on a variable change model rather than the commonly used constant change model. The variable change model is based on the assumption that as a school reaches higher proficiency levels, its proficiency will grow at the pace achieved between 2002 and 2005 by other schools of the same type that attained that higher base level in 2002. Estimating the relationship between the change in proficiency between 2002 and 2005 and the proficiency level attained in 2002 provides a realistic answer to the central analytic question: Are rates of improvement likely to rise, fall, or remain constant relative to current rates?

In addition, results for reading for each subgroup are limited to schools that reported results for that subgroup in each of the four years covered by the data. About 85 percent of the omitted schools had no test takers in a particular subgroup at any time over the four-year period. The remainder had no test takers for one to three years or did not have 10 test takers each year. (Testing results were suppressed in the database if there were fewer than 10 test takers in a group.) About 93.3 percent of schools reported results for whites each year, 81.0 percent reported results for economically disadvantaged students, 68.9 percent for African Americans, 68.4 percent for students with disabilities, 18.0 percent for Hispanics, and 13.6 percent for students with limited English proficiency.

2002 reading and math proficiency levels and 2002–05 trends by school type

In an average school about 74.3 percent of students tested as proficient in reading and math in 2002. The analysis of changes in reading and math proficiency for 2002–05 across all types of schools and separately for elementary, middle, and high schools shows that:

- Between 2002 and 2005 proficiency levels increased in an average school by 6.9 percentage points in reading and 9.8 percentage points in math.
- Average reading proficiency increases were 7.0 percentage points for elementary schools, 7.6 percentage points for middle schools, and 5.9 percentage points for high schools. Average math proficiency increases were 9.2 percentage points for elementary schools, 8.9 percentage points for middle schools, and 13.0 percentage points for high schools.

In an average school in Virginia about 74.3 percent of students tested as proficient in reading and math in 2002

- Across schools of each type there was substantial variation in reading proficiency change between 2002 and 2005. Seventy percent of elementary schools had changes of between –3.3 and +17.3 percentage points, 70 percent of middle schools had changes of between –1.2 and +16.4 percentage points, and 70 percent of high schools had changes of between –3.1 and +14.9 percentage points. Variation in math proficiency was also large. Approximately 70 percent of elementary, middle, and high schools had changes of between –2.0 and +20.4 percentage points.
- Even as proficiency levels rose between 2002 and 2005, the rate of increase consistently declined for all subgroups in all types of schools. For most subgroups and types of schools the decline was about 2.2 percentage points for each 10 percentage point increase in proficiency.

2002 reading proficiency levels and 2002–05 trends by subgroup and school type

The analysis of changes in reading proficiency in six subgroups of students (whites, African Americans, Hispanics, economically disadvantaged students, limited English proficiency students, and students with disabilities) for each type of school showed that:

- In 2002 reading proficiency started at a lower level for all subgroups in middle schools than in elementary and high schools. For five of the six subgroups (all but students with disabilities) proficiency levels were higher in high schools than in elementary schools by 3 to 10 percentage points.
- In 2002 reading proficiency in an average school was highest among white students, at around 80 percent, but the white subgroup had the smallest gains between 2002 and 2005 (3.7 percentage points in elementary schools, 6.1 percentage points in middle schools, and 4.8 percentage points in high schools).

- For Hispanics, African Americans, and economically disadvantaged students in an average school reading proficiency in 2002 was about 61 percent in elementary schools, 54 percent in middle schools, and 71 percent in high schools. Gains were similar for these subgroups in average elementary and middle schools (about 11 percentage points), but in an average high school gains were greatest for Hispanics (7.0 percentage points) and lowest for disadvantaged students (4.5 percentage points).

- For students with limited English proficiency reading proficiency was about 57 percent in an average elementary and high school in 2002 but at only 36 percent in an average middle school. Students with limited English proficiency showed large gains of about 22 percentage points in an average elementary and middle school and smaller gains of about 12 percentage points in an average high school.

- For students with disabilities reading proficiency started at about 54 percent in an average elementary school, 34 percent in an average middle school, and 46 percent in an average high school. Gains averaged 7.4 percentage points in elementary schools, 10.8 percentage points in middle schools, and 14.2 percentage points in high schools.

Projected reading proficiency by subgroup and school type

The variable growth model was used to project steady-state reading proficiency (the proficiency level that will be maintained into the future with only random variations above and below it) for 21 groups—7 groups of students (all students together, plus the six subgroups) in each of the three school types. The key findings are that:

- School proficiency is forecasted to plateau at 90 percent for the white subgroup in an average school; 80 percent for the Hispanic, African American, limited English proficiency, and economically disadvantaged subgroups; and 70 percent for the students with disabilities subgroup.

- For each subgroup 70 percent of schools generally fall within 20 percentage points of the average. For example, in an average elementary school, the African American subgroup plateaus at a 77.5 percent proficiency level, and in 70 percent of elementary schools this subgroup plateaus at between 86.5 percent and 68.5 percent.

- The probability of meeting the 100 percent proficiency target in reading by 2014 is less than 3 percent for 10 of the 21 subgroups, 8–14 percent for 2 subgroups, 29–33 percent for 5 subgroups, and 45 percent or more for 4 subgroups.

The probability of meeting the 100 percent proficiency target in reading by 2014 is less than 3 percent for 10 of the 21 subgroups, 8–14 percent for 2 subgroups, 29–33 percent for 5 subgroups, and 45 percent or more for 4 subgroups

Projected steady-state math proficiency by subgroup and school type

The projections of steady-state math proficiency based on the same model for the same 21 groups yielded the following key findings:

- The forecasted math plateau point for an average school and the distribution of plateau points across individual schools of each type overall and for each of six subgroups were similar to those for reading.

- For an average school math proficiency will plateau at a 95 percent for the white subgroup; 82 percent for the Hispanic, African Americans, limited English proficiency, and economically disadvantaged subgroups; and 74 percent for the students with disabilities subgroup.

WHAT WERE PROFICIENCY LEVELS IN 2002 AND HOW DID THEY CHANGE BETWEEN 2002 AND 2005?

This section of the report examines proficiency levels in reading and math in 2002 and changes each year between 2002 and 2005 for students in the aggregate and in the six subgroups in elementary, middle, and high schools.

Reading and math proficiency by school type

In 2002, the base year for the No Child Left Behind testing program, 74.3 percent of an average school's students tested proficient in reading (figure 1). This level was 13.6 percentage points above the 60.7 percent status (level) standard. The proficiency level in reading in an average school rose to 81.2 percent in 2005, while the standard rose only to 65.0 percent. The average change between 2002 and 2005 was 6.9 percentage points, but there was considerable cross-school variation. About 70 percent of schools had changes between –3.0 and +16.8 percentage points.[1]

In 2002 reading proficiency was 74.2 percent in an average elementary school, 68.8 percent in an average middle school, and 80.1 percent in an average high school. By 2005 these levels had risen to 81.2 percent, 76.4 percent, and 86.0 percent. The average change for elementary schools was 7.0 percentage points, with changes between –3.3 and +17.3 percentage points for about 70 percent of schools. The average change for middle schools was 7.6 percentage points, with changes between –1.2 and +16.4 percentage points for about 70 percent of schools. The average change for high schools was 5.9 percentage points, with changes between –3.1 and +14.9 percentage points for about 70 percent of schools.

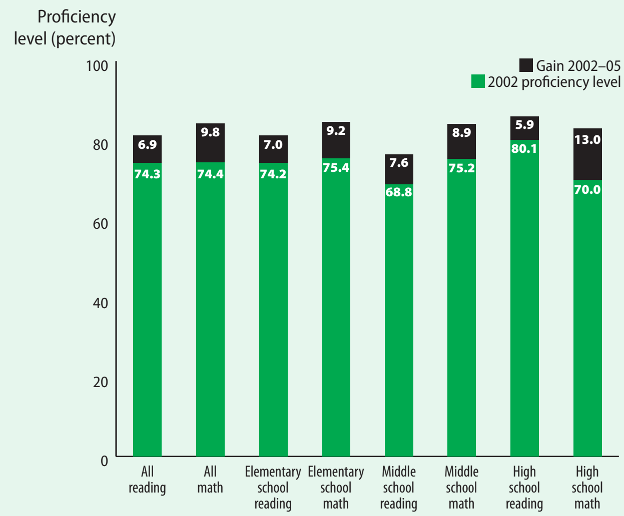
While the average change in the proficiency level in reading in an average school between 2002 and 2005 was 6.9 percentage points, there was considerable cross-school variation: about 70 percent of schools had changes between –3.0 and +16.8 percentage points

In an average school in 2002 the math proficiency level was 74.4 percent. It was 75.4 percent for an average elementary school and 75.2 percent for an average middle school, but 70.0 percent for an average high school (see figure 1). By 2005 proficiency levels had risen about 9 percentage points, on average, in elementary and middle schools. About 70 percent of elementary schools had changes between –2.0 and +20.4 percentage points, and about 70 percent of middle schools had changes between –1.0 and +19.0 percentage points. With an average gain of 13 percentage points, proficiency levels rose more in high schools than in other school types, reaching 83.0 percent and almost closing the gap with elementary schools (84.6 percent) and middle schools (84 percent). About 70 percent of high schools had changes between 3.5 and 22.5 percentage points.

Thus, proficiency rose more in math than in reading, with math proficiency gains outpacing reading gains by 3.0 percentage points. The biggest difference was 7.1 percentage points for high schools. The difference was 2.2 percentage points for elementary schools and 1.3 percentage points for middle schools. The range of variation

FIGURE 1

Average elementary, middle, and high school reading and math proficiency levels for 2002 and proficiency gains for 2002–05



Note: The total statistics cover 1,601 schools; see box 2.

Source: Authors' analysis based on Virginia Department of Education database.

BOX 3

Distribution of students in the six subgroups in the Virginia schools sample

Three of the six subgroups included in the analysis are racial/ethnic groups—whites, African Americans, and Hispanics. Students are assigned to these groups based on information provided by students and parents and sometimes the observation of school officials. Less than 10 percent of students have a race designated as "other" or "not specified." Of the schools in the sample, on average, 63.6 percent of the students are white, 23.9 percent African American, and 3.6 percent Hispanic. Only 23.2 percent of schools report scores for Hispanics, and only about 30 tests are taken by Hispanic students on average in these schools, which is well below the 50 required for the scores of this subgroup to count separately in meeting adequate yearly progress.

The remaining three subgroups are students with limited English proficiency, economically disadvantaged students, and students with disabilities. Limited English proficiency students are placed in that category by school officials who assess students' knowledge of English. Economically disadvantaged students are those who are eligible for free or reduced-price lunch, Temporary Assistance for Needy Families, or Medicaid. Students with disabilities are those who are eligible for services under the Individuals with Disabilities Education Act and who have an Individualized Education Program.[1] Of the schools in the sample, on average, 2.6 percent of the students have limited English proficiency, 22.4 percent are economically disadvantaged, and 13.8 percent have disabilities. Students can enter and leave the economically disadvantaged and disabled groups but can only leave the limited English proficiency group.

Note

1. http://www.doe.virginia.gov/VDOE/Publications/student-coll/06-07/data-elements.xls

in changes across schools was about the same in math proficiency as in reading proficiency, roughly 14.8 percentage points.

Reading proficiency by subgroup

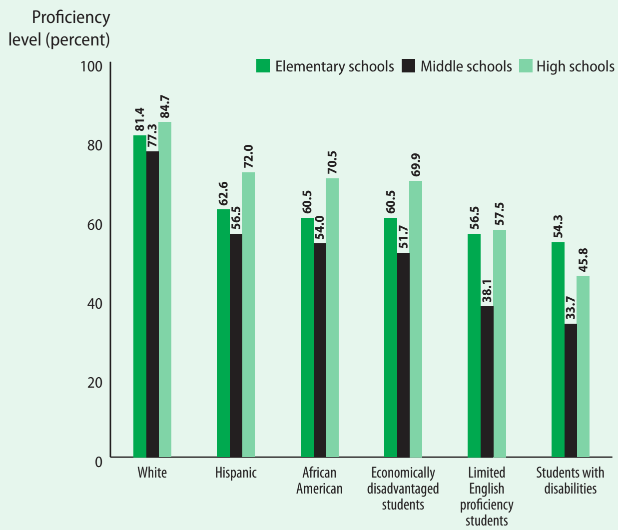
Across all students, reading proficiency levels in an average school increased by 6.9 percentage points over the three years following passage of the No Child Left Behind Act, and math proficiency levels increased by 9.8 percentage points. Because most schools started in 2002 with proficiency levels well above the status standard, annual increases equal to those observed in the early years of the act would be sufficient for roughly 70 percent of schools to reach 100 percent proficiency before 2014. However, some student subgroups started at levels below or near the standard and had lower rates of growth than all students together (see box 3 for a discussion of the six subgroups). The primary focus of the remainder of this report is on reading proficiency. Because proficiency growth was slower in reading than in math, making progress in reading will be the key determinant of whether schools meet the 100 percent proficiency goal by 2014.

Complications of calculating proficiency levels by subgroup. Statistics for all students together always include the full sample of schools. Subgroup statistics only include the schools where proficiency for a given subgroup was reported for each year 2002 through 2005. (Schools were dropped because they had no members of the subgroup or had fewer than 10 members and results were suppressed.) Overall, 93.3 percent of schools reported proficiency for whites for each of the four years, 81.0 percent for economically disadvantaged students, 68.9 percent for African Americans, 68.4 percent for students with disabilities, 18.0 percent for Hispanics, and 13.6 percent for students with limited English proficiency. Also, the same test-taker could be included in several different statistics. For example, a disadvantaged student typically would be in one of the three ethnic groups and could be in each of the two remaining subgroups—students with limited English proficiency and students with disabilities.

Including students whose performance is below average in multiple categories means that a small group of students that is not scoring as proficient could cause a school to fall below the status

FIGURE 2

Average elementary, middle, and high school reading proficiency levels in 2002 by subgroup



Note: Results are displayed from highest to lowest by subgroup.

Source: Authors' analysis based on Virginia Department of Education database.

standard for up to four subgroups. This feature of the No Child Left Behind Act focuses additional attention on raising the performance of students who are members of groups that often have not been the center of attention. It also means that schools with many students in subgroups that typically start off at low proficiency levels will have difficulty meeting absolute level standards, even if the schools excel at raising the performance of those students.

Middle schools usually had the lowest proficiency levels in 2002 and the greatest increases in proficiency. Similarly, the subgroup starting with the highest proficiency levels had small increases in proficiency, while groups starting at especially low levels had especially large gains

Reading proficiency in 2002 by subgroup. In 2002 reading proficiency levels were highest in high schools for each subgroup except students with disabilities, and they were next highest in elementary schools (figure 2).

In each type of school the white subgroup had the highest proficiency level, which was 81.4 percent for an average elementary school, 77.3 percent for an average middle school, and 84.7 percent for an average high school. Hispanics, African Americans, and economically disadvantaged students had similar reading proficiency levels across schools of the same type, but their proficiency levels were 12 percentage points or more lower than those of whites (see figure 2).

In an average school proficiency levels for limited English proficiency students and students with disabilities were below the levels for other subgroups.

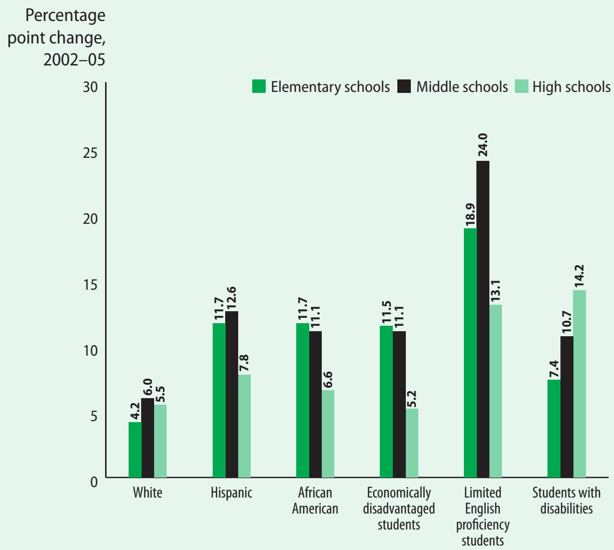
There was substantial variation among schools around the average in all groups. Roughly 70 percent of schools were within 10 percentage points of the mean for the white subgroup, 15 percentage points for Hispanics, 13 percentage points for African Americans and economically disadvantaged students, and 18 percentage points for limited English proficiency students and students with disabilities.

Increases in reading proficiency between 2002 and 2005. Between 2002 and 2005 reading proficiency gains were greatest in middle schools and smallest in high schools, except for students with disabilities, for whom gains were greatest in high schools (see figure 3). For Hispanics, African Americans, and economically disadvantaged students gains were similar in an average elementary school, at about 11.6 percentage points, and in an average middle school, at about 11.7 percentage points. Gains in an average high school were about 5 percentage points lower than in average elementary and middle schools for these subgroups and showed greater variation across the three subgroups. Gains in an average high school were 7.8 percentage points for Hispanics, 6.6 percentage points for African Americans, and 5.2 percentage points for economically disadvantaged students.

For whites gains of 4.2 percentage points in an average elementary school and 6.0 percentage points in an average middle school were 5 or more percentage points less than gains for other subgroups. Gains of 5.5 percentage points for whites in an average high school were also lower than for

FIGURE 3

Percentage point changes in reading proficiency by subgroup and school type, 2002 to 2005



Source: Authors' analysis based on Virginia Department of Education database.

any other subgroup except economically disadvantaged students. For students with disabilities the change was greatest in an average high school (14.2 percentage points), followed by an average middle school (10.7 percentage points) and an average elementary school (7.4 percentage points).

For most subgroups high schools had the highest proficiency levels in 2002 and the smallest increases in reading proficiency. Middle schools usually had the lowest proficiency levels in 2002 and the greatest increases in proficiency. Similarly, the subgroup starting with the highest proficiency levels—whites—had small increases in proficiency, while groups starting at especially low levels—disabled high school students and limited English proficiency students in all types of schools—had especially large gains.

Among individual schools, there was also considerable variation around the mean change in proficiency for each of the 18 groups shown in figure 3. In general, the standard deviations were about equal to the mean. This implies that for elementary school Hispanics about 70 percent of schools were within 12 points of the mean change, or between –2 and +23 points. The main exceptions to the pattern are students with disabilities, for whom standard deviations ranged from 1.6 to 3.0 times the mean, and white elementary school students, for whom the standard deviation was twice the mean.

Estimating the relationship between the change in proficiency between 2002 and 2005 and the proficiency level attained in 2002 provides a realistic answer to the central analytic question: Are rates of improvement likely to rise, fall, or remain constant relative to current rates?

FORECASTING READING AND MATH PROFICIENCY

It is common to assume that the average percentage point change in proficiency will continue to be achieved, even as proficiency levels rise. For this assumption to be true, other things being equal, schools would have to raise the performance of the same number of students above the proficiency threshold each year. But the number of students who test below the proficiency threshold will decline each year, so that a higher fraction of those testing below the proficiency threshold will have to cross that threshold each year. Maintaining a constant proportional increase will be difficult because improved teaching methods are likely to have the greatest impact at the outset, when many students need only small improvements to become proficient. Over time, the students who have not achieved proficiency are likely to require progressively more help to do so.

The observations for Virginia schools for proficiency changes between 2002 and 2005 indicate that the constant improvement assumption does not hold. For each subgroup, average schools' percentage point changes in proficiency between 2002 and 2005 consistently decline as base 2002 proficiency levels increase, suggesting that constant growth models are based on a false assumption and lead to overestimates of future proficiency gains. To obtain more accurate forecasts, simple linear models, which assume that observed trends will continue unchanged, were replaced in this study by a variable growth model that takes into account the evidence that on average percentage point gains between 2002 and 2005 declined as school's 2002 base level of proficiency rose.

Why constant growth models are inappropriate: the change in reading proficiency between 2002 and 2005 by initial proficiency level and school type

Although simple linear (constant growth) models have been used to forecast proficiency level changes (see, for example, MassPartners for Public Schools, 2005), they suffer from three major shortcomings.

First, simple linear models do not take into account the large systematic differences in performance among schools with different initial levels of proficiency. This problem can be overcome by separately forecasting performance for schools with different initial proficiency levels (see, for example, Wiley, Mathis, & Garcia, 2005).

Second, the models do not take into account the range of variation among schools starting at similar proficiency levels. They apply a single number based on the characteristics of a school with average growth and an average starting point. If the variation across schools is small, this defect will not have a large effect, but Virginia schools show tremendous variation in proficiency levels and changes in level.

It is common to assume that the average percentage point change in proficiency will continue to be achieved, even as proficiency levels rise, but the observations for Virginia schools between 2002 and 2005 indicate that the constant improvement assumption does not hold

The author is unaware of any linear forecasts that take into account the range of variation in actual growth rates in forecasting the range of future growth.[2] Rather, linear models provide a single-number forecast and ignore variation around average performance. For example, linear projections often suggest that an average school will meet the 100 percent goal (at least for all students together), but the projections do not indicate what percentage of schools will fall below 100 percent because they start below average or have below average growth rates.

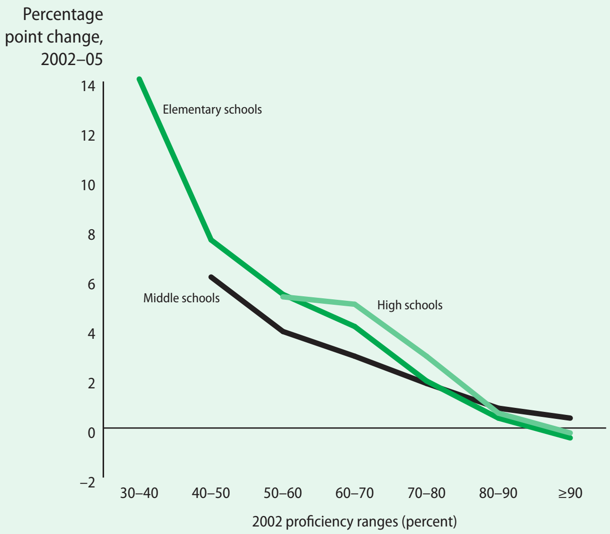
More broadly, the No Child Left Behind Act specifies that states should use appropriate statistical techniques to determine whether a school has failed to make adequate yearly progress in a given year. Virginia is among the minority of states that do not apply a statistical approach that takes into account year-to-year variation in a school's performance in comparing a school's proficiency level in one year with that year's status standard. However, few, if any, states apply a statistical approach that adequately distinguishes long-term trends from large transitory fluctuations in looking at the change in proficiency over time—what statisticians call distinguishing signal from noise.

A few academic papers have examined the signal-to-noise issue. Using North Carolina school-level data similar to the data used here, Kane & Staiger (2002a, b) conclude that about 75 percent of year-to-year changes in proficiency level reflect random variation. Because these fluctuations are often large, data for several years need to be combined to determine how proficiency changes over time. Kane and Staiger show that the correlation of the change in school-level proficiency from one year to the next averages –0.35. This implies that schools showing increases in one year have about a 50 percent chance of showing a decrease in the next year and vice versa. This result implies that the application of the "safe harbor" provision to one-year changes will often falsely identify random variation as true improvement and will often fail to identify true improvement. It strongly reinforces the decision applied in this study to average growth over three years as the basis for forecasting future growth.

The third problem with a linear projection model is that it explicitly assumes that percentage point increases in proficiency for a given subgroup at a given type of school will remain constant over time. Conceptually, this assumption is questionable since it requires that equal numbers of

FIGURE 4

Average annual change in reading proficiency from 2002 to 2005 for schools with different 2002 proficiency levels, by school type



Note: Schools of each type were grouped into 10 percentage point "bins" based on 2002 proficiency levels, and then the average change between 2002 and 2005 was calculated for the schools in each bin.

Source: Authors' analysis based on Virginia Department of Education database.

students become proficient each year even as the level of proficiency rises and the pool of nonproficient students shrinks. More commonly it is observed that at best a constant proportion of students in the nonproficient pool becomes proficient, causing the rate of improvement to decrease as the level increases.

Figure 4 illustrates why the assumptions underlying constant growth models are inconsistent with the patterns observed in the Virginia database. The figure shows the average annual change in proficiency between 2002 and 2005 for each type of school for schools starting in 2002 at different proficiency levels. Schools of each type are grouped into 10 percentage point "bins" based on 2002 proficiency levels, and the average change for 2002 to 2005 is calculated for the schools in each bin. For example, elementary schools with proficiency levels of 60–70 percent in 2002 experienced an average annual change of 4.1 percentage points between 2002 and 2005.[3] Growth in proficiency declines at higher 2002 base proficiency levels, with the data points falling in a downward-slanting line. For elementary and high schools annual change declines by 2.2 percentage points for each 10 percentage point increase in starting proficiency levels, while for middle schools change declines by 1.2 percentage points.

Most important, figure 4 shows that the average change was between 0.4 and 0.8 percentage point across the three school types for schools whose 2002 proficiency levels were 80–90 percent. This is roughly 4–10 times less than the average change across school types whose 2002 proficiency levels were 60–70 percent. For schools with 2002 proficiency above 90 percent, the change across types ranged from –0.4 to +0.4 percentage point. The contrast between schools with initial proficiency levels above 90 percent and those with proficiency levels below 60 percent is especially important since it provides strong evidence that the assumption used in the constant average change models is untenable—schools attaining high levels of proficiency do not sustain increases near the average gain achieved by all schools together and do not come close to achieving the percentage point gains reached by schools with low levels of proficiency in 2002.[4]

The pattern shown in figure 4 held for 20 of the 21 cases examined (the seven population groups—all students together and each of the six subgroups in each of three types of schools). The exception was in the second-smallest sample, in which the reliability of the data was questionable.

Why a variable change model is more appropriate: estimating the relationship between reading proficiency level and growth

A variable growth model—which empirically estimates how annual changes in reading proficiency systematically vary as levels of proficiency rise—was developed and estimated based on the relationship between annual changes in reading proficiency between 2002 and 2005 in Virginia schools and the level of proficiency in 2002. The estimate draws on a formulation that has been used before (see Jacobson, LaLonde, and Sullivan, 2004 and 1993). The basic model is:

(1) $$\Delta P_{2002-05} = \alpha + \beta P_{2002} + \varepsilon$$

where P_{2002} is a school's proficiency level in 2002, and $\Delta P_{2002-05}$, the change in proficiency, equals the three-year average annual change in proficiency 2002–05 ($\Delta P_{2002-05} = (P_{2005} - P_{2002}) / 3$.

An attractive feature of this model is that it forecasts the long-term steady-state point (*S*) that schools will reach with the simple calculation shown in equation 2.

(2) $$S = \alpha / \beta$$

where α is the intercept coefficient in equation 1, and β is the slope coefficient. This feature of the model greatly reduces the confidence interval surrounding the point estimate, whereas in other models predictions *n* years in the future require multiplying the model's coefficients *n* times and consequently multiplying the error term by *n*.

This model is similar to value-added models recently approved for use as part of the U.S. Department of Education's No Child Left Behind Growth Model Pilot Program (see, for example, Tennessee Department of Education, 2006). Both models have an autoregressive form (future test results are predicted based on observed test results, usually with no other variables included), and both models take confidence intervals surrounding point estimates into account. As a result, they produce statistically sound estimates that take random variation into account in estimating systematic trends (see, for example, Wright, Sanders, and Rivers, 2005).

To predict proficiency through 2014 by accurately capturing how proficiency is likely to change as proficiency levels increase, the variable change model estimates the relationship between the change in proficiency between 2002 and 2005 and the level of proficiency in 2002 for each subgroup and type of school

However, the value-added models are applied to student-level, not school-level, data; they are applied to test scores, not proficiency levels; and they generally are used to predict growth three years into the future, not over the much longer period needed to reach steady-state points.[5] They also require sophisticated data management systems that link data for the same student in successive years. Most states, including Virginia, did not have the capacity to do this during the period studied. Moreover, while Virginia and many other states are putting such systems into place, data for at least three years will have to be accumulated before value-added estimates can be produced.

To predict proficiency through 2014 by accurately capturing how proficiency is likely to change as proficiency levels increase, the variable change model estimates the relationship between the change in reading proficiency between 2002 and 2005 and the level of proficiency in 2002 for each subgroup and type of school. Ordinary least squares regressions are used to estimate equation 1.

The following equations were estimated using data for all students together for 1,018 elementary schools, 298 middle schools, and 287 high schools. Standard errors are presented below the coefficients.[6]

For elementary schools:

(3) Change = 16.4 + [−0.189 × Proficiency level] adjusted R^2 =.543.
(0.41) (0.005).

For middle schools:

(4) Change = 10.3 + [−0.114 × Proficiency level] adjusted R^2 =.308.
(0.70) (0.010).

For high schools:

(5) Change = 19.6 + [−0.221 × Proficiency level] adjusted R^2 =.52.6
(1.00) (0.012).

Equation 3 predicts that elementary schools with 30 percent proficiency in one year would raise proficiency by 10.7 percentage points, on average, the following year (16.4 + [–0.189 × 30] = 10.7). However, on average, schools with a 60 percent proficiency level in one year would increase proficiency by 5.1 points in the next year (16.4 - 0.189 × 60 = 5.1). This is because equation 3 indicates that, on average, a 1 percentage point increase in proficiency in one year, decreases the improvement in the next year by 0.189 percentage point.[7]

Finally, the adjusted R^2 measures how close the data points are to the regression line specified by equation 3. In this case the regression line explains 54.3 percent of the variation in the elementary school data. Coupled with the standard errors being about one-fortieth of the coefficients, these statistics indicate that the model "fits" the data unusually well.

The coefficients in equation 5 for middle schools are about half those for elementary schools. However, the standard errors for the coefficients are about twice as large as those for elementary schools. This is in keeping with the fact that there are only about one-fourth as many middle schools as elementary schools in Virginia. The adjusted R^2 is substantially lower, indicating that 30.8 percent of the variation is explained. Thus, the fit of the middle school regression line is not nearly as good as the fit of the elementary school regression. Equation 5 shows that on average middle schools at the 60 percent proficiency level in any one year will show a 3.5 percentage point gain the next. It also shows that 95 percent of the middle schools at the 60 percent level will show gains of between 1.1 and 5.9 percentage points.

The coefficients in equation 6 for high schools are similar to those for elementary schools, as is the adjusted R^2—52.6 percent of the variation is explained. However, the standard errors surrounding the coefficients are a bit greater than those for middle schools. Equation 6 shows that on average high schools at the 60 percent proficiency level in one year will show a 6.4 percentage point gain the next and that 95 percent of those high schools will show gains of between 0.2 and 12.6 percentage points.

Estimation of steady-state proficiency levels and time-paths for reaching those levels

The equations in the preceding section for all students in each type of school (and in appendix B for students in each subgroup in each type of school) can be used to estimate "steady-state" points—the levels of proficiency that, once reached, will be maintained into the future with only random variations above and below them. These steady-state points are based on the key assumption that as schools increase their proficiency they will grow at the successively declining rate achieved by schools that were at that higher base level in 2002. This section describes the mean and variance of the estimated steady-state points. The next section presents additional evidence bearing on the accuracy of the key assumption.

The more a school's proficiency level is below some fixed point, the faster the model predicts that its proficiency level will grow, but growth slows and eventually stops as proficiency rises

There are steady-state points because the more a school's proficiency level is below some fixed point, the faster the model predicts that its proficiency level will grow, but growth slows and eventually stops as proficiency rises. Moreover, if a school's proficiency level rises above the fixed steady-state point because of random fluctuations, the model predicts that the proficiency level will fall back toward the steady-state point. For example, for schools with proficiency levels above 90 percent in 2002, proficiency declined about 0.1 to 0.4 percentage point on average per year through 2005 depending on school type. In short, the model is consistent with the evidence produced by researchers such as Kane & Staiger (2002a) that there are large fluctuations around central tendencies and that central tendencies can be discerned by observing behavior over a series of years.

The steady-state level implied by the model can be derived by solving equation 1 for the proficiency level that generates a change of zero. As shown in equation 6, the steady-state level equals negative-one times the intercept coefficient divided by the slope coefficient.

(6) Steady-state proficiency level = (–1 × Intercept coefficient / Slope coefficient).

To illustrate the steady-state calculation equation 6 is applied to the coefficients for elementary schools shown in equation 3. The steady-state point for elementary schools is 86.8 percent (–16.4 /–0.19). About 70 percent of elementary schools are predicted to reach steady-state proficiency levels of between 81.1 percent and 90.7 percent. About 95 percent of elementary schools are predicted to reach steady-state proficiency levels of between 76.6 percent and 95.1 percent.

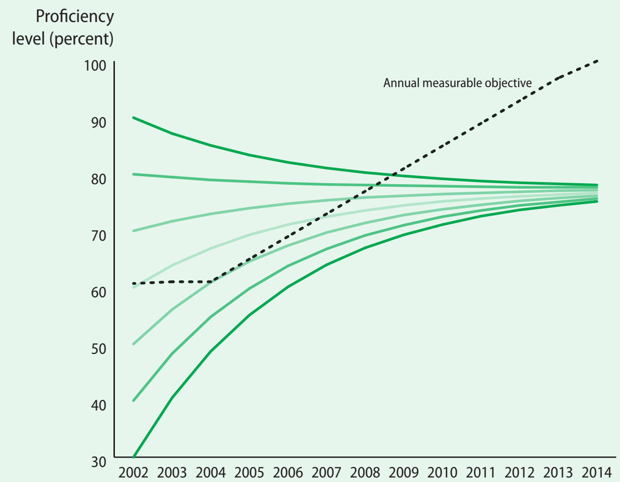
Thus, the model predicts that most elementary schools will eventually sustain proficiency levels of within about 9 percentage points of 87 percent. While the model predicts that there will be some variability in the expected steady-state point across schools, less than 1 percent of elementary schools are likely to sustain proficiency levels near the 100 percent goal for 2014.

The model predicts that most elementary schools will eventually sustain proficiency levels of within about 9 percentage points of 87 percent

Figure 5 shows a typical time-path for reaching the steady-state point for groups of schools where 2002 proficiency for a given subgroup in a given school type was 30, 40, 50, 60, 70, 80, and 90 percent. In this illustration the initial proficiency level for a given subgroup and school type averaged across all schools is 62 percent, and the steady-state level is 78 percent. The dashed black line shows the path of the reading proficiency standard as it reaches 100 percent in 2014. Each of the time-paths is estimated by forecasting reading proficiency from 2002 to 2014 using the variable change model for schools starting at different proficiency levels in 2002.

The forecasts are derived by inserting a school's assumed starting point for a given subgroup into equation 1 and estimating the equation using that subgroup's proficiency level in 2002 and the change in proficiency over 2002–05 for each school for which there was sufficient data for that subgroup. This produces an estimate of the change to be added to the starting level to forecast the next year's proficiency level. The next year's level is then plugged into the model to estimate the following year's level, and that gain is added to the next year's level, and the process is repeated to estimate proficiency levels for each year 2006 through 2014.

FIGURE 5
Forecasts of reading proficiency from 2002 to 2014 using the variable change model for schools starting at different proficiency levels in 2002



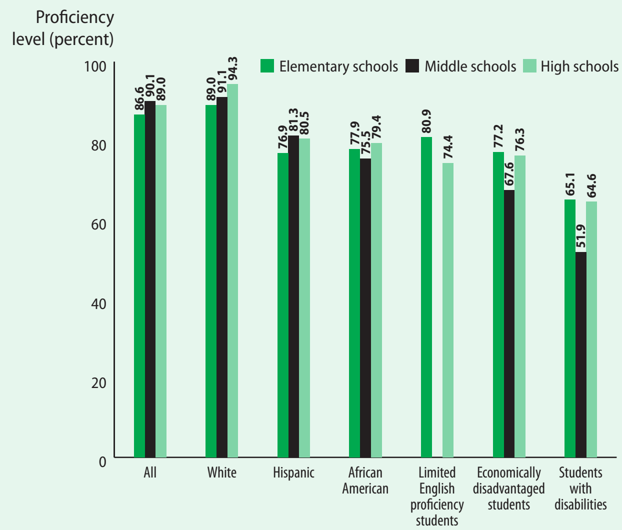
Note: The bowed lines show the estimated time-path of proficiency for schools starting at different levels. These lines converge at a proficiency level of about 78 percent, the steady-state point.

Equation 1 is very similar when estimated for each of four subgroups—Hispanics, African Americans, economically disadvantaged students, and limited English proficiency students—when estimated separately for elementary, middle, and high schools. Figure 5 takes one of these equations and displays the time-paths for schools in one subgroup in one type of school with different proficiency levels in 2002. The shapes of the paths are also similar for whites and students with learning disabilities, but the steady-state point is considerably higher for whites and substantially lower for students with disabilities.

There are substantial increases in proficiency over the first six years (2002–07) for groups of schools starting out with much lower than average proficiency levels (2002 proficiency levels of 30, 40, and 50 percent), but annual increases decline dramatically as these subgroups approach the steady-state point. Although subgroups starting out at high proficiency levels (80 and 90 percent) show declines, they exhibit a similar pattern of larger movements toward the steady-state point over the first six years, followed by slower movements subsequently.

Figure 5 shows that by 2010 four school groups—those with 2002 proficiency of 30, 40, 50, and 60 percent—will attain a proficiency level of above 70 percent on average, but their levels will be below the 81 percent proficiency level standard for that year. The school groups starting at 30, 40, and 50 percent proficiency never exceed the standard, even though their absolute increases are especially large. The group of schools starting at a 60 percent proficiency level do not exceed the standard in the period when projections are needed, but exceeded the standard through 2005. Perhaps most important, the four subgroups that exceed the initial 60.4 percent standard by 10 percentage points or more—those with 2002 proficiency levels of 70, 80, and 90 percent—all fall below the standard by the end of 2008.

FIGURE 6
Steady-state reading proficiency levels by subgroup and school type



Note: Results are displayed from highest to lowest by subgroup. Middle school limited English proficiency students are not included because the standard errors of the regression coefficients were so large that the coefficients were meaningless.

Source: Authors' analysis based on Virginia Department of Education database.

Steady-state reading proficiency levels by subgroup and school type

Figure 6 displays the average steady-state reading proficiency levels for each school type and student subgroup. (Results are not displayed for middle school limited English proficiency students because the data from the unusually small number of schools in this sample contained several outliers, which led to unrealistic estimates.)

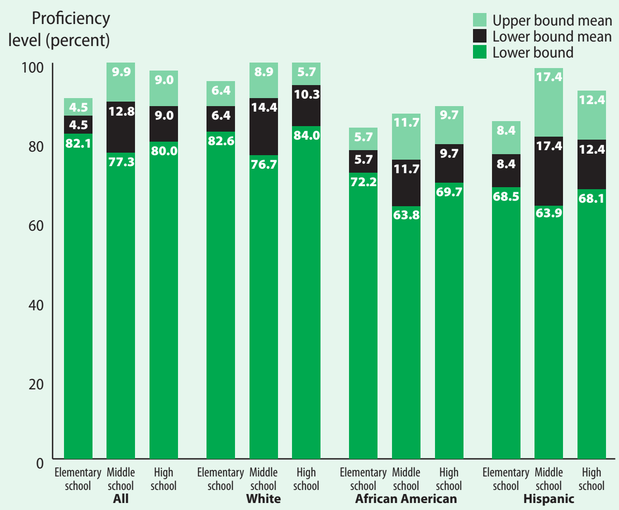
Equations of the form of equation 1 were estimated, and these equations were then used to calculate the steady-state reading proficiency levels by dividing the intercepts by the slopes, as shown in equation 6. (Appendix B displays the equations estimated using the change data along with relevant goodness-of-fit statistics.)

The steady-state reading proficiency levels are similar for a given subgroup across the three types of schools with the exception of the economically disadvantaged students and students with disabilities subgroups, where middle school steady-state points are about 10 percentage points below average. Steady-state levels for Hispanics are about 79.7 percent in all types of schools. Levels for African Americans and limited English proficiency students are about 77.7 percent, slightly lower than for Hispanics, but about the same across school types. Levels for whites range from 89.0 percent in elementary schools to 94.3 percent in high schools.

Levels for economically disadvantaged students are about 76.8 percent in elementary and high schools, but only 67.6 percent in middle schools. These levels are about the same as for African

FIGURE 7

Upper and lower bounds of the confidence interval for steady-state reading proficiency levels for whites, African Americans, Hispanics, and all students, by school type

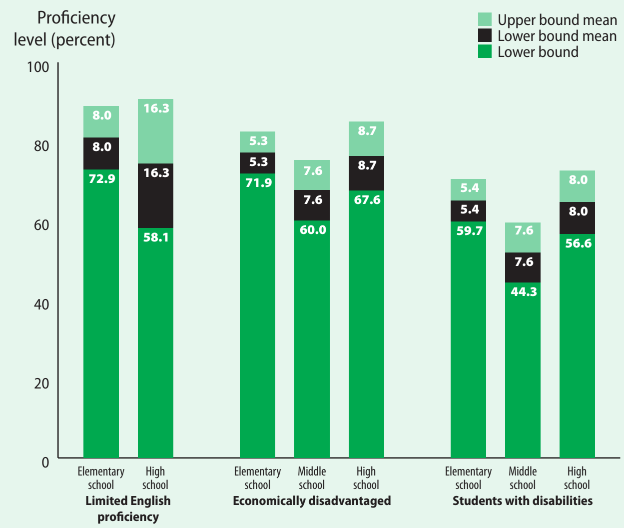


Note: The height of the bottom two sections of the bars combined indicates the mean estimate for steady-state reading proficiency, and the height of the three sections together indicates the upper bound estimate.

Source: Authors' analysis based on Virginia Department of Education database.

FIGURE 8

Upper and lower bounds of the confidence interval for the steady-state reading proficiency levels for limited English proficiency students, students with disabilities, and economically disadvantaged students subgroups, by school type



Note: The height of the bottom two sections of the bars combined indicates the mean estimate for steady-state reading proficiency, and the height of the three sections together indicates the upper bound estimate.

Source: Authors' analysis based on Virginia Department of Education database.

Americans and Hispanics in elementary schools, 3.1 percentage points below the level for African Americans in high schools, and 7.9 percentage points below that for African Americans in middle schools.

Finally, steady-state levels for students with disabilities are about 64.8 percent in elementary and high schools but 51.9 percent in middle schools.

Figures 6 displays point estimates of the average steady-state proficiency level for schools in each subgroup. The model also predicts variations in the steady-state level for schools of a given subgroup. Figures 7 and 8 display the confidence interval (range of variation) around the point estimates using one-standard-deviation values. For example, 70 percent of elementary schools will attain steady-state proficiency levels for whites of between 82.8 percent and 95.5 percent, values that are 6.4 percentage points below and above the predicted mean of 89.2 percent.

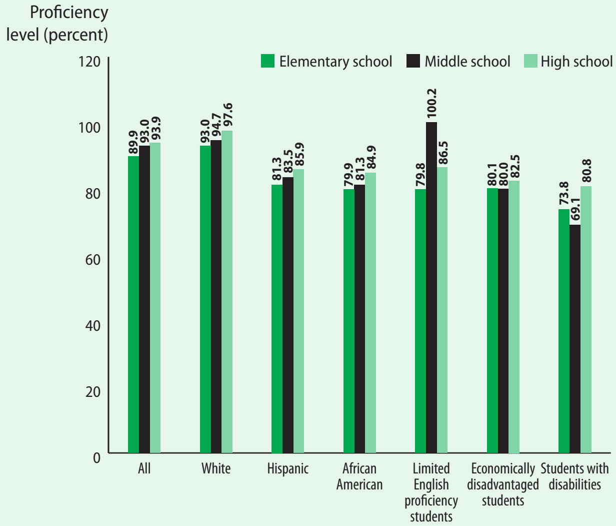
On average the range of variation across the 20 groups in figures 7 and 8 is 18.3 percentage points. However, the range of variation is especially large, about 33.7 points, for Hispanics in middle school and limited English proficiency students in high school—both groups that include small numbers of schools. On average, the range of variation is about twice as great in middle school subgroups as in elementary school subgroups, and about 60 percent greater in high schools than in elementary schools. These differences result in large part because there are about 3.5 times more elementary schools in the sample than middle schools or high schools. (For all students together there are 1,018 elementary schools, 298 middle schools, and 287 high schools in the sample.)

Steady-state math proficiency levels by subgroup and school type

Figure 9 displays the average steady-state math proficiency levels for the same set of schools for

FIGURE 9

Steady-state math proficiency levels by subgroup and school type



Source: Authors' analysis based on Virginia Department of Education database.

which steady-state reading proficiency levels were displayed in figure 6, based on the same estimating procedures. The pattern of results across subgroups and school types is similar.

The primary difference between steady-state math and steady-state reading proficiency levels is that the math levels are higher. In general, the differences are proportional to the higher growth rates observed for math than for reading, as shown in figure 1. Math levels are 3.3 percentage points higher on average for the five elementary school groups with the smallest standard errors (all students together, whites, Hispanics, African Americans, and economically disadvantaged students). The math levels average about 5 percentage points higher for the same five groups for middle and high schools.

For students with disabilities the steady-state proficiency levels are higher for math than for reading levels by about 9 percentage points in elementary schools and by about 16 percentage points in middle and high schools. These differences probably reflect a greater effect of student disabilities on reading than on math. High school limited English proficiency students' math levels are higher by about 12 percentage points, reflecting the fact that achieving proficiency on a math test is much less dependent on English fluency than is achieving proficiency on a comprehensive reading test.

African American and economically disadvantaged middle school students also showed substantially higher steady-state proficiency levels in math than in reading. The difference was 8 percentage points for African Americans and 12 percentage points for disadvantaged students. Differences persisted in high school but were smaller, at 6.5 percentage points for African Americans and 6.2 percentage points for economically disadvantaged students.

Because in almost every other way the math results were similar to the reading results, the details are not presented here. Appendix B describes the regression equations used to estimate the math steady-state points.

Past performance as a guide to future performance

The projections of steady-state proficiency levels provide indicators of the extent to which Virginia schools will be able to reach a central No Child Left Behind goal of having every student proficient in reading and math by 2014. The results are consistent with the patterns observed in the data: it is evident that annual proficiency growth rates for all students together and for individual subgroups for 2002–05 are much lower in schools that started at high base proficiency levels in 2002 than in schools that started at low levels. Also, schools in which subgroups started at an 85 percent proficiency level in 2002 rarely showed any subsequent growth. Rather, they tended to decline by small amounts.

However, the estimates do not precisely reflect how observed trends and future performance might diverge from past performance. To examine potential sources of inaccuracy, several additional analyses were carried out (details are reported in appendix C). These efforts focused on three possibilities:

- The estimated model does not perfectly fit the patterns observed in the cross-sectional database. To test for this, alternative functional forms were estimated for the change-level relationship.

- There are factors that influence 2002–05 growth rates other than 2002 proficiency levels, factors that are not included in the model. This was tested by checking the effect of adding variables to the basic linear model.

- There are factors that will influence growth-level relationships beyond 2005 that cannot be observed in the period studied. To test for this, future changes that might influence proficiency growth-level relationships were considered.

The conclusion from these tests is that it cannot be ruled out that large discontinuities with the past will affect the steady-state levels that Virginia schools achieve. However, such changes are not visible on the horizon, with one possible exception—major changes in adequate yearly progress standards that may be enacted as part of the No Child Left Behind reauthorization.

The overall conclusion with respect to the accuracy of the estimated growth model is that, on balance, the model appears to provide reasonably accurate estimates of the steady-state level at which the performance of each subgroup will eventually plateau in each type of school and the way actual steady-state points are likely to vary from these estimates. In other words the steady-state point estimates and confidence intervals appear statistically sound.

On balance, the model appears to provide reasonably accurate estimates of the steady-state level at which the performance of each subgroup will eventually plateau in each type of school and the way actual steady-state points are likely to vary from these estimates

At the same time, additional research is warranted to provide more information about how differences across schools and districts in factors other than proficiency levels will affect steady-state proficiency levels and growth rates. For example, analysis of factors such as district enrollment, concentrations of economically disadvantaged students in a school, and the pace of improvement for schools with below-average performance, holding observable factors constant, could provide information useful for making policy decisions.

CONCLUSIONS ABOUT PROFICIENCY LEVELS AND METHODOLOGY

This report reaches conclusions about both the findings and the methodology.

Findings about proficiency levels and actual and predicted changes in proficiency levels

Virginia public schools made major strides in fulfilling the goals of the No Child Left Behind Act from 2002 through 2005:

- Math and reading proficiency levels showed average gains of 7 to 10 percentage points.

- Increases tended to be highest among schools and subgroups that had the lowest initial proficiency levels. Limited English proficiency students, who started at the lowest proficiency level (45 percent), had the largest increases.

Virginia public schools are likely to continue to make progress in meeting key No Child Left Behind goals in the near term:

- Realistic estimates of future progress indicate that proficiency levels will be about 4 percentage points higher on average in 2008 than in 2005.

- These increases will be largest for schools and subgroups with below-average proficiency levels.

However, the rate of progress will slow as schools and subgroups exceed an 80 percent proficiency level, and few schools will be able to consistently achieve proficiency levels close to 100 percent, as required by the No Child Left Behind Act.

As a result, sooner or later, the proficiency levels of virtually all Virginia schools for most subgroups will fall below the status standard as the standard rises from 69 percent in 2006 to 100 percent in 2014. It becomes progressively harder to increase proficiency levels as proficiency levels rise. Indeed, schools with proficiency levels higher than 90 percent in 2002 experienced reductions of about 0.2 percentage point on average by 2005.

In 2006 about 22 percent of Virginia schools were labeled in need of improvement. Most of these schools did not meet the status standard for one or more groups and also failed to meet the safe harbor standard. The percentage of schools failing to meet the status standard will increase over the next few years—assuming that proficiency standards and the rigor of the testing program do not materially change—as the status standard increases by 4 percentage points a year. However, the percentage of schools that will be labeled in need of improvement because they do not meet adequate yearly progress standards for two or more successive years is difficult to estimate because many schools will be able to meet the safe harbor standard.

Meeting the safe harbor standard is likely to become increasingly common as that standard falls from an increase of 2 points to 0 points as proficiency levels rise from 80 percent toward 100 percent. Moreover, because the status standard increases by a constant 4 percentage points a year, it is certain that the safe harbor standard will determine whether schools meet adequate yearly progress for most, if not all, subgroups. A preliminary investigation (not reported here) suggests that over the next two years 30–50 percent of Virginia's schools will fail the status standard for one or more groups, but will make adequate yearly progress based on the safe harbor standard.

Advantages and implications of using the variable change model for forecasting proficiency improvements

The review of the literature found that previous efforts to forecast improvements in proficiency levels used some type of constant growth model. Such models are based on the assumption that a group's observed increase over a base period will continue unchanged into the future. These forecasts were also based on the average change for a group of schools.

The rate of progress will slow as schools and subgroups exceed an 80 percent proficiency level, and few schools will be able to consistently achieve proficiency levels close to 100 percent

The methodology used in this study forecasts proficiency using a variable change model in which the observed changes are within one standard deviation of the average change—the range of variation that includes the 70 percent of schools with changes closest to the mean. This extension provides a more realistic estimate of the percentage of schools in any group likely to reach a target proficiency level. This is especially important for a state like Virginia, where the standard deviation around the mean is close to 10 percentage points.

The variable change model adjusted predicted change as a function of the proficiency level attained in a given year. This was based on strong evidence that the higher the proficiency level in base year 2002, the smaller the increase in proficiency attained by 2005. The level-change relationship turned out to be well described by a linear model of the form:

$$\text{Proficiency change} = \text{Intercept coefficient} + \text{Slope coefficient} \times \text{Proficiency level}$$

Taking the level-change relationship into account yields estimates that proficiency levels will plateau at a steady-state point below the 100 percent target for 2014. Because the approach used here takes into account much more information about how proficiency changes as levels increase and the range of variation in proficiency changes across schools than do models in the literature, it represents an important advance.

Further, the model developed here is easy to apply to the school-level data commonly available, makes it easy to estimate steady-state levels, and uses regression analysis to produce statistics that describe how well the model fits the data as well as variations around the mean values of starting points and change amounts.

A more subtle advantage of the model is that it minimizes the statistical uncertainty surrounding the estimates because it directly estimates steady-state levels by dividing the intercept coefficient by the slope coefficient. (Technically, it produces narrow confidence intervals.) In contrast, constant change models repeatedly use the same statistic to estimate changes from one year to the next, thereby multiplying the effect of any statistical error and increasing the confidence intervals surrounding point estimates.

The model developed here is easy to apply to the school-level data commonly available

Unclear from the work completed so far is whether the change-level relationship in other states will fit the linear model as closely as it did in Virginia. However, the model will soon be applied to Kentucky data, and researchers in other regions are invited to apply the model to states in their regions to test its general usefulness.

Another important methodological finding is that producing subgroup estimates is essential to identifying potential problems in meeting adequate yearly progress goals. It was difficult to determine trends in year-to-year fluctuations using the Virginia data covering only four years, because the fluctuations are large and do not follow a clear pattern. Thus, averaging proficiency change over a four-year period appeared to produce better indicators of long-term trends than using information about change over one or two years.

This result is important for accurately forecasting proficiency and has significant implications for the validity of standards that are based on year-to-year changes, such as the safe harbor standard. Change-based standards generate high pass rates when year-to-year fluctuations in proficiency levels in individual schools are large, as they are in Virginia, even when they are not correlated with rising long-term trends. For example, a school that reaches 90 percent proficiency in year one, falls to 86 percent proficiency in year two, and rises to 88 percent proficiency in year three will meet the safe harbor standard in year 3 without showing any evidence of long-term growth.

Additional uses for the model

Models such as the one developed here have been used in estimating changes in individual student test scores, but as far as the author knows, have not been used to estimate growth using school-level data. The model has applications beyond forecasting proficiency levels. It can determine how a given school's increase in proficiency compares with that of other schools with about the same level of proficiency in a base year. For example, for schools with initial proficiency levels of 60–70 percent, the model can estimate what changes will occur in the top 15 percent of schools, schools whose performance is above average but below that of the top 15 percent, schools whose performance is below average but above that of the bottom 15 percent, and schools in the bottom 15 percent.[8]

These estimates provide a means of identifying the schools that are doing far better than their peers and the schools that should be capable of showing substantial improvement. This information could help achieve key No Child Left Behind goals by ensuring that resources and remedial actions are focused on schools that are most clearly underperforming and that have the best chance of improving performance by employing techniques that worked well in schools facing similar challenges. Schools that are performing well relative to their peers could potentially do better, but it is much more difficult to figure out how to boost their performance.

The model developed here can also help assess the strengths and weaknesses of different standards for judging performance. For example, the model can rank a school according to how far its progress is above or below average in improving performance (relative to schools starting out at similar proficiency levels). An alternative standard, such as the No Child Left Behind status standard or safe harbor standard, can then be used to establish a second set of ranks for the same group of schools. The two rankings can be compared to assess the extent to which the alternative standards distinguish underperforming schools from schools that are performing well, given the challenges they face.

Investigating the soundness of the safe harbor standard is especially important. This standard will become the major determinant of "making" adequate yearly progress as both proficiency levels and the status standard rise above 80 percent. This is because the safe harbor standard requires an annual decrease of at least 10 percent in students testing below the proficiency level. For a school at 80 percent proficiency, that translates into a required increase of only 2 percentage points, whereas the status standard rises by 4 percentage points a year.

However, this analysis and the earlier analysis of North Carolina data by Kane & Staiger, 2002a, show that it is very difficult to discern long-term trends in proficiency growth based on comparing levels from one year with the next. Because school proficiency levels show substantial volatility from year to year, many schools can meet the safe harbor standard without making any long-term progress in proficiency. Kane & Staiger, 2002b, discuss how to resolve this problem using North Carolina data. Acquiring Virginia data back to 1999 would provide a long enough time series to replicate and update the Kane & Staiger analysis.

Averaging proficiency change over a four-year period appeared to produce better indicators of long-term trends than using information about change over one or two years

Finally, the analysis has looked only at the key No Child Left Behind requirement to boost scores above the proficiency threshold, which is equivalent to a test score of about 65 percent. The same techniques could be used to analyze increases in the percentage of students with scores above the advanced threshold, which is equivalent to a test score of about 85 percent. This analysis would provide information about the effect of the No Child Left Behind Act on students who become proficient and students who start out well above proficient. It would also shed light on the extent to which schools that outperform their peers in increasing the percentage of students who achieve proficiency also outperform their peers in increasing the percentage who pass the advanced threshold.

APPENDIX A
VARIABLES IN DATABASE USED TO ESTIMATE PROFICIENCY CHANGE-LEVEL MODELS

For each school:

1. District
2. Name
3. Lowest grade
4. Highest grade
5. Region of district
6. Number of schools in the district
7. Number of high schools in the district
8. School type (elementary, middle, high, mixed)
9. County name
10. City name (if any)
11. Metropolitan statistical area name (if any)

Number of test-takers and percent proficient for math and reading for each year 2002–2005 for:

1. All students
2. Whites
3. African Americans
4. Hispanics
5. Limited English proficiency students
6. Economically disadvantaged students
7. Students with disabilities

Separate regressions are run for elementary, middle, and high schools for each of the seven population subgroups.

Additional right-hand-side school variables:

1. Total number of test-takes is a measure of school size
2. Distributions of students in each group

Additional right-hand-side district variables:

1. Total number of test-takes (measure of district size)
2. Number of elementary, middle, and high schools in district
3. Distribution of students in each group (measure of diversity)

County-level census variables (matched to each district):

1. Land area of county (square miles)
2. Population in 2000
3. Population growth 1990–2000
4. Percent of households in poverty 1999
5. Percent of population over 25 with high school degree
6. Percent of population over 25 with bachelor's degree
7. Population density (people per square mile)
8. Population ages 5–17
9. Public school enrollment 1999
10. Change in enrollment 1989–99
11. Median household income 1997
12. Crime rate 1999
13. Federal funds and grants per capita 1999
14. Change in federal funds per capita 1989–99
15. Name of MSA containing county, if any

Common core school-level variables

1. Number of students in the third grade
2. Number of students in the fifth grade
3. Number of students in the eighth grade

Common core district-level variables

1. Total staff
2. Student/teacher ratio
3. Full-time equivalent teachers
4. Total general revenue
5. Percent of revenue from local sources
6. Percent of revenue from state sources
7. Percent of revenue from federal sources
8. Fall membership (number of students)
9. Local revenue per student
10. State revenue per student
11. Federal revenue per student
12. Total expenditures
13. Capital outlays
14. Instructional expenditures
15. Supplemental service expenditures
16. Other expenditures
17. Total nonelective expenditures
18. Total elective expenditures

APPENDIX B
REGRESSIONS USED TO ESTIMATE PROFICIENCY CHANGE-LEVEL RELATIONSHIPS AND STEADY-STATE LEVELS, BY SCHOOL TYPE AND SUBGROUP

TABLE B1

Regressions used to estimate reading proficiency change-level relationships and steady-state levels

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Elementary schools							
All							
Intercept	16.4	0.411	39.9	86.6	4.5	0.543	1,018
Proficiency 2002	–0.189	0.005	–34.8				
White							
Intercept	16.4	0.584	28.0	89.0	6.4	0.416	936
Proficiency 2002	–0.184	0.007	–25.8				
African American							
Intercept	17.5	0.596	29.4	77.9	5.7	0.457	655
Proficiency 2002	–0.225	0.010	–23.5				
Hispanic							
Intercept	20.9	1.112	18.8	76.9	8.4	0.581	181
Proficiency 2002	–0.273	0.017	–15.8				
Limited English proficiency							
Intercept	20.9	0.930	22.4	80.9	8.0	0.630	159
Proficiency 2002	–0.258	0.016	–16.4				
Economically disadvantaged							
Intercept	17.7	0.563	31.4	77.2	5.3	0.437	824
Proficiency 2002	–0.229	0.009	–25.3				
Students with disabilities							
Intercept	14.8	0.607	24.4	65.1	5.4	0.428	627
Proficiency 2002	–0.228	0.011	–21.6				

(continued)

TABLE B1 (CONTINUED)

Regressions used to estimate reading proficiency change-level relationships and steady-state levels

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Middle schools							
All							
Intercept	10.3	0.698	14.8	90.1	12.8	0.308	296
Proficiency 2002	–0.114	0.010	–11.5				
White							
Intercept	13.2	1.052	12.6	91.1	14.4	0.289	282
Proficiency 2002	–0.145	0.014	–10.7				
African American							
Intercept	12.9	0.933	13.9	75.5	11.7	0.299	242
Proficiency 2002	–0.171	0.017	–10.2				
Hispanic							
Intercept	13.7	1.404	9.8	81.3	17.4	0.471	55
Proficiency 2002	–0.169	0.024	–7.0				
Limited English proficiency							
Intercept	11.3	1.493	7.6	131.0	50.5	0.126	35
Proficiency 2002	–0.086	0.035	–2.4				
Economically disadvantaged							
Intercept	15.5	0.808	19.2	67.9	7.6	0.462	262
Proficiency 2002	–0.228	0.015	–15.0				
Students with disabilities							
Intercept	10.2	0.680	14.9	51.9	7.6	0.296	269
Proficiency 2002	–0.196	0.018	–10.7				

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
High schools							
All							
Intercept	19.6	0.998	19.7	89.0	9.0	0.526	287
Proficiency 2002	−0.221	0.012	–17.8				
White							
Intercept	17.9	0.988	18.1	94.3	10.3	0.493	276
Proficiency 2002	−0.190	0.012	–16.4				
African American							
Intercept	19.5	1.212	16.1	79.4	9.7	0.506	206
Proficiency 2002	−0.246	0.017	–14.5				
Hispanic							
Intercept	24.0	1.910	12.6	80.5	12.4	0.718	52
Proficiency 2002	−0.298	0.026	–11.4				
Limited English proficiency							
Intercept	18.8	2.100	8.9	74.4	16.3	0.704	23
Proficiency 2002	−0.252	0.035	–7.3				
Economically disadvantaged							
Intercept	20.8	1.225	17.0	76.3	8.7	0.544	211
Proficiency 2002	−0.273	0.017	–15.9				
Students with disabilities							
Intercept	16.3	0.932	17.4	64.6	8.0	0.469	199
Proficiency 2002	−0.252	0.019	–13.3				

(continued)

TABLE B2

Regressions used to estimate math proficiency change-level relationships and steady-state level

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state math level: Point estimate	Steady-state math level: Standard deviation	Adjusted R^2	Number of observations	Difference between math and reading steady-state level
Elementary schools								
All								
Intercept	19.0	0.425	44.6	89.9	4.5	0.587	1,018	3.3
Proficiency 2002	–0.211	0.006	–38.0					
White								
Intercept	19.4	0.545	35.7	93.0	5.7	0.518	936	4.0
Proficiency 2002	–0.209	0.007	–31.7					
African American								
Intercept	17.8	0.642	27.8	81.3	7.0	0.414	655	3.4
Proficiency 2002	–0.219	0.010	–21.5					
Hispanic								
Intercept	21.5	1.204	17.8	79.9	10.6	0.549	180	3.0
Proficiency 2002	–0.269	0.018	–14.8					
Limited English proficiency								
Intercept	22.9	1.111	20.6	79.8	9.2	0.643	158	–1.1
Proficiency 2002	–0.286	0.017	–16.8					
Economically disadvantaged								
Intercept	19.7	0.624	31.6	80.1	5.9	0.437	824	2.9
Proficiency 2002	–0.246	0.010	–25.3					
Students with disabilities								
Intercept	15.0	0.563	26.6	73.8	6.8	0.396	611	8.7
Proficiency 2002	–0.203	0.010	–20.0					

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state math level: Point estimate	Steady-state math level: Standard deviation	Adjusted R^2	Number of observations	Difference between math and reading steady-state level
Middle schools								
All								
Intercept	15.5	0.658	23.6	93.0	9.2	0.560	297	2.1
Proficiency 2002	–0.167	0.009	–19.5					
White								
Intercept	17.7	0.930	19.1	94.7	11.3	0.500	282	3.6
Proficiency 2002	–0.187	0.011	–16.8					
African American								
Intercept	16.1	0.829	19.5	83.5	10.9	0.460	242	8.0
Proficiency 2002	–0.193	0.013	–14.4					
Hispanic								
Intercept	20.6	2.108	9.8	81.3	20.8	0.554	54	0.0
Proficiency 2002	–0.253	0.031	–8.2					
Limited English proficiency								
Intercept	9.0	2.447	3.7	100.2	121.0	0.119	35	–31.0
Proficiency 2002	–0.090	0.038	–2.4					
Economically disadvantaged								
Intercept	17.4	0.756	23.0	80.0	8.8	0.517	262	12.1
Proficiency 2002	–0.217	0.013	–16.7					
Students with disabilities								
Intercept	13.3	0.708	18.8	69.1	11.0	0.310	266	17.2
Proficiency 2002	–0.192	0.018	–10.9					

(continued)

TABLE B2 (CONTINUED)

Regressions used to estimate math proficiency change-level relationships and steady-state level

School type and variable name	Coefficient	Standard error	*t*-statistic	Steady-state math level: Point estimate	Steady-state math level: Standard deviation	Adjusted R^2	Number of observations	Difference between math and reading steady-state level
High schools								
All								
Intercept	17.0	0.689	24.7	93.9	9.3	0.550	287	4.9
Proficiency 2002	−0.181	0.010	−18.7					
White								
Intercept	16.0	0.937	17.1	97.6	14.1	0.392	276	3.3
Proficiency 2002	−0.164	0.012	−13.4					
African American								
Intercept	16.2	0.905	17.9	85.9	13.0	0.413	206	6.5
Proficiency 2002	−0.189	0.016	−12.1					
Hispanic								
Intercept	18.6	2.282	8.2	84.9	27.9	0.443	52	4.4
Proficiency 2002	−0.219	0.034	−6.4					
Limited English proficiency								
Intercept	17.4	3.064	5.7	86.5	42.9	0.488	23	12.1
Proficiency 2002	−0.201	0.043	−4.7					
Economically disadvantaged								
Intercept	18.6	1.237	15.0	82.5	14.0	0.377	211	6.2
Proficiency 2002	−0.226	0.020	−11.3					
Students with disabilities								
Intercept	13.2	0.878	15.0	80.8	17.2	0.258	196	16.2
Proficiency 2002	−0.163	0.020	−8.3					

APPENDIX C
PAST PERFORMANCE AS A GUIDE TO FUTURE PERFORMANCE

The forecasts of the extent to which Virginia schools will be able to reach the central No Child Left Behind goal of having every student proficient in reading and math by 2014 are consistent with the patterns observed in the data. It is evident that growth rates between 2002 and 2005 are much lower in schools with high proficiency levels in 2002 for all students together and for individual subgroups than in those that started with low proficiency levels. Also, schools in which groups reached 85 percent proficiency in 2002 rarely showed any subsequent growth. Rather, they tended to decline by small amounts.

However, the estimates do not precisely reflect how observed trends and future performance might diverge from past performance. This section describes several analyses conducted to examine potential sources of inaccuracy. These analyses focus on three possibilities:

- The estimated linear change-level model does not perfectly fit the patterns observed in the cross-sectional database.
- There are factors other than 2002 proficiency levels that influence 2002–05 growth rates and that are not included in the model.
- There are factors that will influence growth-level relationships beyond 2005 that cannot be observed in the period studied.

Tests of alternative functional forms for the change-level relationship

One test used to examine how well a linear change-level model fits the data is to introduce the square of the 2000 proficiency level into equation 1. Including the square term only slightly improved the fit of the equations and did not materially change the estimates of the steady-state levels. This test suggests that the "true" level-change relationship is very close to linear.

A more flexible functional form was also tested using a piecewise linear model (see appendix D). As shown in equation C1, separate slope and intercept terms were estimated for schools with 2002 proficiency levels below 65 percent, between 65 and 80 percent, and above 80 percent in 2002.[9]

(C1) Proficiency change =
Intercept coefficient – 1
+ Slope coefficient – 1
× Proficiency level (given proficiency <65%)
+ Intercept coefficient – 2
+ Slope coefficient – 2
× Proficiency level (given proficiency 65%–80%)
+ Intercept coefficient – 3
+ Slope coefficient – 3
× Proficiency level (given proficiency >80%).

Dividing each of the three intercept coefficients by the corresponding slope coefficients produced separate estimates of the steady-state point for schools starting in 2002 at each of the three proficiency ranges. The differences in the steady-state levels across the three proficiency groups and between each of the three groups and the estimate using a single intercept and slope coefficient are indicators of the extent to which the change-level relationship can be described by a single straight line and of the differences in the steady-state levels of schools with different starting points.

Table C1 displays the four steady-state estimates for all students together in elementary schools, along with the standard errors of the estimates and the adjusted R^2 for the two models. The basic model steady-state estimate is almost identical to the piecewise model steady-state estimates for the two groups of schools starting at or above a 65 percent proficiency level. However, the standard errors of the piecewise estimates are more than 10 times greater than those of the basic estimate. This is the case even though the R^2, a measure of how closely the data fit the model, is about 20 percent greater for the piecewise linear model.

The similarity of the steady-state estimates, coupled with the very large standard errors, suggests

TABLE C1

Comparisons between the basic linear model and the piecewise linear model for elementary schools

		Piecewise linear model		
	Basic linear model	Proficiency level <65%	Proficiency level 65%–80%	Proficiency level >80%
Steady-state point	86.6	78.4	86.0	87.7
Standard error	4.5	18.7	50.9	126.9
Number of observations	1,018	231	407	380
Adjusted R^2	0.543	0.697		

that the basic linear model produces estimates superior to those of the piecewise linear model.[10] The large standard errors surrounding the piecewise linear point estimates are a natural outgrowth of that model, making predictions based on a narrow band of results and, in some cases, making predictions well outside the proficiency range covered by the data.

The substantially lower estimated steady-state proficiency level for schools below a 65 percent proficiency level than for schools at higher levels suggests that schools performing at low levels in 2002 will have difficulty catching up with schools performing at higher levels. The differences in the standard errors suggest that there is considerably more uniformity in the change-level relationship among schools in the less than 65 percent group than among schools in the 65–80 percent group and more uniformity in the change-level relationship among schools in the 65–80 percent group than among schools in the more than 80 percent group.

The lower steady-state estimate for schools below 65 percent proficiency might be due to those schools having more students in the subgroups with lower estimated steady-state levels. To test this hypothesis, results from the basic model were separately compared with those from the piecewise model for each subgroup in each type of school. Table C2 shows that the piecewise model's steady-state estimates are usually lower than the basic model's estimates, but in most cases the differences in the point estimates are less than 2 percentage points.[11] These small differences for estimates disaggregated by subgroup confirm that much of the far larger differences observed for all students together were due to not taking into account differences in the distribution of the subgroups across each of the three piecewise groupings. The differences were negative because the basic model produced slightly lower steady-state estimates. This is an important finding because it suggests that the basic model estimates are slight overestimates of the steady-state rates derived from assuming that the observed change-level relationships will apply beyond 2005.

TABLE C2

Average difference in steady-state level estimates between the basic model and the piecewise model for subgroups by school type

Proficiency level	Elementary school	Middle school	High school
<65%	–0.1	–5.6	1.7
65%–80%	–1.1	–0.6	–3.5
>80%	–2.5	5.6	–1.9

Note: The figures represent the piecewise linear estimate minus the basic model estimates. Thus, negative numbers indicate that the piecewise linear estimates are lower than the basic model estimates.

The effect of adding variables to the basic linear model

The piecewise linear results indicate that not controlling for student characteristics produces misleading estimates of the steady-state level when proficiency is examined for all students together. In this section the analysis is extended to assess the extent to which not taking other factors into account affects the accuracy of steady-state estimates. To examine the effect of other factors, a database was assembled that included the 59 variables described in appendix A for the 1,602 schools and 131 districts in the sample. These factors include school-level student characteristics such as the proportion of students from economically disadvantaged (low-income) families, county-level variables such as median income and percent of the population over age 25 with a bachelor's degree, and district-level variables describing factors such as total enrollment and the amount of education revenue from federal, state, and local sources. The effect of these variables was then systematically examined on the key regression parameters in equation 1 used to compute steady-state levels by adding those variables to the change in level equations.

Table C3 displays the 11 additional variables that had more than a trivial effect on the overall explanatory power of the regressions when added to the basic model described by equation 1. Most of these variables were either statistically significant or strongly affected the statistical significance of other variables based on t-tests.[12] The regression estimates showed that reading proficiency in 2002, percentage of economically disadvantaged students, and number of students in the district were statistically significant for all three school types. Two other variables were statistically significant for elementary schools alone: a dummy variable set to one if there were 50 or more economically disadvantaged students in a given school and federal funding per student in the district.

The regression coefficient was also multiplied by each variable's standard deviation to estimate the impact of a one-standard-deviation change in a given variable on the change in proficiency between 2002 and 2005. Table C3 shows the average results for all three school types combined. The variables are listed from strongest to weakest effect. The table should be interpreted with great care, as the variables are not necessarily mutually independent. To the extent that the variables are correlated with one another, the rankings will not reflect the strength of a given characteristic. The regression results for the number of students in the district and the population density (population per square mile) are most likely to capture "true" independent effects.

The key result in table C3 is that a one-standard-deviation shift in reading proficiency (the only variable in the basic model) has about five times

TABLE C3

The effect of selected variables on proficiency growth, 2002–05

Variable name	Standard deviation of variable	Coefficient times standard deviation
Reading proficiency 2002	11.8	–2.57**
Percentage economically disadvantaged	20.6	–0.63**
Number of students in district	44,877	0.47**
Percentage limited English proficiency	6.3	–0.28
More than 50 economically disadvantaged students	0.442	–0.18*
Percentage African American	26.6	–0.12
Percentage students with disabilities	7.0	0.12
Population per square mile in district	1,548	0.11
Federal funding per student in district ($)	265	0.10*
Number of test-takers	111	0.07
Percentage Hispanic	7.3	–0.06
Expenditures per student in district ($)	1,237	0.03

*Statistically significant at the 95 percent level for elementary schools alone.

**Statistically significant at the 95 percent level for elementary, middle, and high schools.

Note: Unless otherwise noted the variables are school specific.

the effect of a one-standard-deviation shift in the percentage of economically disadvantaged students, the added variable with the strongest effect on the change in proficiency. This result suggests that omitted variables are unlikely to have much effect on the steady-state estimates—a result strongly reinforced by tests described below.

The number of students in the district is in third place overall in the strength of its effect, and it is also statistically significant for all school types. The percentage of limited English proficiency students has the next largest effect but is not statistically significant. This insignificance most likely stems from the small number of schools (about 100 out of 1,602) having 25 or more limited English proficiency test-takers. For elementary schools alone, having 50 more economically disadvantaged students had a relatively large and statistically significant effect.

Federal funds per student in the district had a small but statistically significant effect for elementary schools. A one-standard-deviation increase in funding per student ($265 with a mean of $563) was associated with a 0.35 percentage point increase in proficiency in elementary schools. This represents about a 15 percent increase in the average change, which was 2.34 percentage points. The effect of federal spending was essentially zero for middle and high schools.

Overall, table C3 provides an indication of which variables in the database might have an important influence on proficiency growth, but with the possible exception of district size and population density, it might not tell us much about the magnitude of the effect because of intercorrelation among the included variables. The variables not included in the table had small effects that were not statistically significant. Thus, it is hard to see how they could influence the accuracy of estimates derived from equation 1.

Information in table C3 about the strength of the effects indicates which factors are likely (and unlikely) to be important. To obtain more definitive information about how the introduction of the additional variables affected both the overall explanatory power of the model and the steady-state proficiency estimates, the R^2 of the equations was examined and steady-state levels were computed using equation 2.

Table C4 shows how R^2 varies across the different models for each type of school.[13] Adding the 11 additional variables to the basic model increased the explanatory power (R^2) by only 1.8 percentage points for elementary schools, while using the 11 variables without the 2002 proficiency level explained only 24.6 percent of the variation in the 2002–05 change in proficiency. Adding the 11 variables increased the R^2 by 10.0 percentage points for middle and high schools, but using them without the 2002 proficiency level explains only about 4 percent of the variation in the 2002–05 change in proficiency.

A third model included separate intercept and slope coefficients for schools that were in districts with above and below average enrollment. These regressions were run to more clearly describe how large a difference this key variable makes in the estimation of steady-state levels. In this case the R^2 increases by about 14 percentage points for elementary and high schools and by about 35 percentage points for middle schools. These results suggest that increases in proficiency differ substantially across districts with above and below average enrollment and that the differences are about three times greater for middle schools than for elementary and high schools.

Table C4 also shows how the steady-state estimates vary when different models are used. When the 11 additional variables shown in table C3 are added to the basic model, the high school estimates hardly change. They show an increase of just 0.8 percentage point. However, adding the 11 additional variables lowers the estimates for elementary schools by 3.4 percentage points and for middle schools by 10.5 percentage points.

In each case, the steady-state level equals one minus the intercept coefficient divided by the slope coefficient. Thus, the size of the reduction in the

TABLE C4
The effect of different specifications on the explanatory power of change regressions and on the estimated steady-state levels

	Elementary schools	Middle schools	High schools
Adjusted R-squared			
Basic model	0.543	0.308	0.526
Basic + 11 variables	0.561	0.407	0.626
11 variables	0.246	0.043	0.033
Basic model above and below average district size	0.692	0.654	0.664
Steady-state estimates			
Basic model	86.6	90.1	89.0
Basic + 11 variables	83.2	79.6	89.8
Basic model district-size divisions			
Below average	85.0	86.2	88.3
Above average	88.9	95.6	90.1

steady-state estimate stemming from adding the additional 11 variables is a measure of how much the coefficients change. It does not imply anything about how steady-state estimates vary across schools with different characteristics.

To address the question of how steady-state levels are affected by the 11 additional variables, separate intercept and slope coefficients were included for schools in which the value of a given variable is either above or below average. (Basically, the data used in the regression were split into two groups based on whether the value of an individual variable was above or below average for a given school.) Table C4 illustrates how the splitting system works for district enrollment—a variable with a large and statistically significant effect on proficiency change from 2002 to 2005.

The last two lines of table C4 indicate that large districts consistently reach higher steady-state levels than do small districts. However, the difference is only 1.8 percentage points for high schools, about twice as large for elementary schools (3.9 percentage points), and more than twice as large again for middle schools (9.4 percentage points).

Overall, these results suggest that adding additional variables to the basic model alters only slightly the powerful relationship between proficiency in 2002 and the change in proficiency between 2002 and 2005. Put another way, there is a relatively narrow range of variation in steady-state levels across schools with substantially different characteristics. In particular, there is little reason to believe that taking additional factors into account would modify the basic conclusion that few schools will come close to reaching 100 percent proficiency for all students together or for any subgroup.

On the other hand, further analysis of the variation associated with the few statistically significant variables with large effects could reveal differences in proficiency growth that have important implications for policy. For example, other things equal, it is plausible that larger districts would have more success in raising proficiency than smaller districts would, because large districts can take advantage of economies of scale. If confirmed, this hypothesis might suggest that small districts need additional assistance identifying and implementing steps to improve proficiency.

Future changes that might influence proficiency level-growth relationships

The first section examined the assumption that the level-growth relationship is linear, as specified in equation 2. The second section examined the assumption that proficiency levels are the primary determinant of the changes in proficiency, as specified in equation 2, and that other factors do not have a large effect on proficiency growth that might adversely affect the accuracy of the steady-state estimates.

This section examines the possibility that even if the basic model is appropriately specified for estimating the level-growth relationships observed in 2002–05 there will be changes in the future that affect the accuracy of these estimates. In contrast to the first two sections, where the accuracy of the assumptions can be empirically tested, there is no way to be sure whether past trends will continue into the future. Rather, factors that might cause future level-growth relationships to depart from those in the past can be identified, and the likelihood that these departures will be large can be examined.

It is important to recognize that the variable growth model does not assume that there will be no systematic improvements in the education system due to factors such as advancements in curricula, teaching methods, and teacher quality or that there will be no changes in the composition of student populations. Rather, the assumption is that the rate of change that occurred between 2002 and 2005 will continue into the future. Thus, deficiencies in the model are most likely to stem from substantial discontinuities with the past.

While some exceptionally potent innovation in education or a major demographic change in Virginia cannot be ruled out, such changes usually evolve at a pace that rarely accelerates or decelerates substantially. Major changes in the development and diffusion of education innovations or in demographic trends are not visible on the horizon. Virginia's education accountability system and governance systems are relatively mature, and the current performance of Virginia public schools is among the best in the country.

One discontinuity that could arise is if Virginia substantially altered its testing procedures and tests. One major change is that starting in 2007 Virginia will use its existing Standards of Learning science tests to determine adequate yearly progress under the No Child Left Behind Act. Also, Virginia changed its procedures in 2006 to test each grade 3 through 8, rather than only grades 3, 5, and 8. It also altered the reading and math tests to place more emphasis on skills learned each year and less on cumulative skill acquisition. Virginia tried to minimize the effect of these changes by designing the new tests to be comparable with the old ones in rigor and in scores required to achieve proficiency. Nevertheless, the new procedures could lower proficiency levels in the short term. It is common for teachers and supervisors to require time to adjust to changes in accountability systems.

A higher probability threat to the accuracy of the steady-state estimates is that the impact of the accountability system could increase. Requiring testing in grades 4, 6, and 7 will increase the number of test-takers in elementary schools by 50 or 100 percent and in middle schools by 100 or 200 percent, depending on whether the elementary-middle school break is after the fifth or sixth grade. This will increase the probability that individual subgroups will have 50 or more test-takers and therefore count separately in making adequate yearly progress.

In addition, incentives to boost performance could increase substantially as the 2007 status standard of 73 percent rises by 4 percentage points a year, because an increasing number of relatively high performing schools will fall below the status standard, and incentives to boost proficiency are strong only when levels are near or below those needed to meet adequate yearly progress.

If incentives matter, it is possible that steady-state levels are understated for schools with proficiency levels above 80 percent in 2002. It is also possible that incentives will weaken for some schools and subgroups. For example, as the status standard rises, local educators charged with improving the performance of schools that are persistently falling further and further below the standard could become discouraged as they perceive that trying harder is futile.

Thus, the strongest incentives to improve performance under adequate yearly progress may apply to schools that are just below the status standard, where a little more effort is most likely to make a major difference in a school's rating.

There is a modest amount of literature on the incentive effects of high-stakes educational testing. For example, Ladd (2004) provides evidence that teachers in North Carolina responded to a bonus plan by transferring to schools likely to qualify for the bonus. Figlio & Lucas (2004) show that home buyers are sensitive to test scores in neighborhood schools. However, authoritative results could not be found on the empirical link between adequate yearly progress provisions and test scores or on how incentive effects vary in relation to how far a school's performance is above or below a given standard. This is surprising, because such a demonstration would be important evidence in deciding how accountability systems should be altered when No Child Left Behind is reauthorized. What is clear is that schools in two states, Florida and North Carolina, have made major improvements in both state and national test scores after adopting rigorous accountability systems as well as a host of other initiatives to improve performance. What is less clear is the precise source of these improvements (Figlio & Rouse, 2005).

In the absence of clear-cut evidence the structure of No Child Left Behind accountability mandates was examined to assess how strong the incentives are to improve performance and whether their effects could be examined empirically. Two conclusions were reached about the extent to which incentives are likely to change. The first is that the status standard in Virginia will recede in importance as the level of the status standard rises. This is because schools can make adequate yearly progress based on meeting whichever standard creates a lower hurdle, and the safe harbor standard presents a lower hurdle than the status standard. For subgroups in schools where proficiency is 73 percent in 2007, exactly at the 2007 status standard, increasing proficiency sufficiently to meet the safe harbor standard—10 percent of the percentage that are not proficient—will require gains that decline from 2.7 percentage points to 0 as proficiency increases. In contrast, average annual increases of 4 percentage points are required for these subgroups to meet the status standard, which rises by 4 percentage points a year. For subgroups in schools at 85 percent proficiency in 2007 the safe harbor standard will fall uniformly from requiring an increase of 1.5 percentage points to 0 as proficiency increases. Meeting the status standard will require an average increase of just over 2.0 percentage points each year. (This is because the gap between the 100 percent target for 2014 and the 2007 level is 15 percentage points and schools have seven years to meet the target: 15/7 = 2.14). The second is that the year-to-year volatility in a school's proficiency scores is so great that most schools above 80 percent proficiency will meet adequate yearly progress requirements based on safe harbor provisions at least once every two years, whether their proficiency levels show a long-term rising trend or not.

Therefore, the rising status standards are not likely to create strong incentives to improve performance, and the strength of future accountability systems may depend largely on whether the safe harbor standard is revised or some other type of growth-based standard is introduced as part of the No Child Left Behind reauthorization process. Revising the standards is a possibility because the academic literature suggests that the safe harbor standard does not produce statistically meaningful indications that schools are improving proficiency in the long run (see Kane & Staiger, 2002b).[14]

It might be possible to conduct an empirical test of the strength of incentive effects because some schools have fewer than 50 test-takers in a given subgroup and are thus exempt from separately counting that group in making adequate yearly progress, while others have 50 or more test-takers and are not exempt. To determine whether an empirical test might be feasible, the number of schools in each subgroup, the number with 50 or more test-takers in each year between 2002 and 2005, and the number with fewer than 50 test-takers each year were calculated. (Omitted are schools that sometimes have 50 or more test-takers.)

There are 51 or fewer schools with enough Hispanics, limited English proficiency students, and students with disabilities to have these subgroups count separately every year (table C5). There are 31 schools with too few students as a whole to apply the regular adequate yearly progress standards; they are treated as special cases by Virginia. There also are 173 schools with too few white students to have this group count separately in making adequate yearly progress every year.

Analysis suggests that a minimum of about 100 schools are needed in both the count and no-count groups to discern modest differences between the two groups. Further, there should be enough schools to permit separate tests for elementary, middle, and high schools. Thus, there are only two subgroups for which a basic test can be conducted—African Americans and economically disadvantaged students. The analysis looked at whether proficiency growth among the disadvantaged students subgroup was greater in schools where disadvantaged subgroups "counted." The results are ambiguous for several reasons.

First, schools with a high proportion of economically disadvantaged students had slower proficiency growth rates for this group than schools with small percentages, holding proficiency levels in 2002 constant. Presumably, this is because it is easier to raise the proficiency levels of members of this group when they are relatively few in number. As a result, the negative effects of differences in difficulty may have concealed any positive incentive effects. Second, several factors may have weakened incentives among schools where the disadvantaged students subgroup's performance was separately counted for making adequate yearly progress. For example, incentives could be weak if

TABLE C5

Number and percentages of schools with no data for each subgroup and number with 50 or more test-takers each year during 2002–05, by subgroup

	Subgroups with too few uncovered schools				Subgroups with too few covered schools		
	All	White	African American	Economically disadvantaged students	Hispanic	Limited English proficiency	Students with disabilities
Number of schools with four years of data	1,602	1,495	1,104	1,298	289	218	1,096
Percentage of schools with four years of data	100.0	93.3	68.9	81.0	18.0	13.6	68.4
Number of schools with 50 or more test-takers each year, 2002–05	1,524	1,163	485	473	*44*	*22*	*51*
Percentage of schools with 50 or more test-takers each year, 2002–05	95.1	77.8	43.9	36.4	15.2	10.1	4.7
Number of schools with less than 50 test-takers each year, 2002–05	*31*	173	434	486	183	99	867
Percentage of schools with less than 50 test-takers each year, 2002–05	1.9	11.6	39.3	37.4	63.3	45.4	79.1

Note: The term *covered* means that a school has 50 or more test takers in a given subgroup each year from 2002 through 2005, so that school is held separately responsible for meeting adequate yearly progress for that subgroup in each year. The term *uncovered* means that a school did not have 50 or more test takers. Numbers in italics indicate cases where there are too few schools to test incentive effects for a given subgroup. To test incentives there should be at least 100 covered and 100 uncovered schools of given type.

that subgroup's performance was already considerably above the status standard or if such large improvements were needed to meet status or safe harbor standards that the schools believed that the barriers were too difficult to overcome.

Finally, there are nonacademic factors, not described in the database, that influence whether a school meets adequate yearly progress. These factors include teacher quality, safety, the percentage of students tested, and high school attendance and graduation rates. In some cases where economically disadvantaged students counted separately for adequate yearly progress, school officials knew that they would not meet adequate yearly progress requirements for nonacademic reasons.

APPENDIX D
PIECEWISE LINEAR REGRESSIONS USED TO ESTIMATE READING PROFICIENCY CHANGE-LEVEL RELATIONSHIPS AND STEADY-STATE LEVELS, BY SCHOOL TYPE AND SUBGROUP

TABLE D1

Elementary schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
All							
Proficiency level 2002: 74.2							
Change in proficiency 2002–05: 2.34							
Piecewise model							
Schools with 2002 level <65							
Intercept	20.5	1.0	20.2	78.4	18.7	0.697	231
Level 02	–0.262	0.018	–14.4				
Schools with 2002 level 65–80							
Intercept	15.4	2.0	7.9	86.0	50.9		407
Level 02	–0.179	0.027	–6.7				
Schools with 2002 level >80							
Intercept	8.2	2.2	3.7	87.7	126.9		380
Level 02	–0.094	0.025	–3.7				
Basic model							
Intercept	16.4	0.4	39.9	86.6	4.5	0.543	1,018
Level 02	–0.189	0.005	–34.8				
White							
Proficiency level 2002: 81.4							
Change in proficiency 2002–05: 1.41							
Piecewise model							
Schools with 2002 level <65							
Intercept	14.7	2.9	5.1	100.1	183.4	0.524	71
Level 02	–0.147	0.049	–3.0				
Schools with 2002 level 65–80							
Intercept	18.4	2.3	7.8	86.2	51.2		297
Level 02	–0.213	0.032	–6.7				
Schools with 2002 level >80							
Intercept	13.9	1.7	8.3	89.3	45.0		568
Level 02	–0.156	0.019	–8.2				
Basic model							
Intercept	16.4	0.6	28.0	89.0	6.4	0.416	936
Level 02	–0.184	0.007	–25.8				

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
African American							
Proficiency level 2002: 60.5							
Change in proficiency 2002–05: 3.91							
Piecewise model							
Schools with 2002 level <65							
Intercept	17.5	1.0	17.7	77.9	22.3	0.666	394
Level 02	–0.224	0.019	–11.7				
Schools with 2002 level 65–80							
Intercept	19.2	4.3	4.5	77.4	90.9		202
Level 02	–0.249	0.060	–4.2				
Schools with 2002 level >80							
Intercept	9.7	8.0	1.2	71.6	–266.3		59
Level 02	–0.136	0.093	–1.5				
Basic model							
Intercept	17.5	0.6	29.4	77.9	5.7	0.457	655
Level 02	–0.225	0.010	–23.5				
Hispanic							
Proficiency level 2002: 62.6							
Change in proficiency 2002–05: 3.88							
Piecewise model							
Schools with 2002 level <65							
Intercept	22.1	1.8	12.0	75.7	32.0	0.711	105
Level 02	–0.291	0.035	–8.4				
Schools with 2002 level 65–80							
Intercept	8.5	8.2	1.0	78.1	–94.7		52
Level 02	–0.109	0.114	–1.0				
Schools with 2002 level >80							
Intercept	23.9	10.9	2.2	80.7	421.8		24
Level 02	–0.296	0.123	–2.4				
Basic model							
Intercept	20.9	1.1	18.8	76.9	8.4	0.581	181
Level 02	–0.273	0.017	–15.8				

(continued)

TABLE D1 (CONTINUED)

Elementary schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Economically disadvantaged							
Proficiency level 2002: 60.5							
Change in proficiency 2002–05: 3.82							
Piecewise model							
Schools with 2002 level <65							
Intercept	19.1	0.9	21.1	73.8	16.7	0.656	522
Level 02	–0.259	0.017	–15.1				
Schools with 2002 level 65–80							
Intercept	22.6	4.0	5.7	76.9	63.6		240
Level 02	–0.294	0.055	–5.3				
Schools with 2002 level >80							
Intercept	9.4	10.2	0.9	72.2	–128.4		62
Level 02	–0.130	0.119	–1.1				
Basic model							
Intercept	17.7	0.6	31.4	77.2	5.3	0.437	824
Level 02	–0.229	0.009	–25.3				
Students with disabilities							
Proficiency level 2002: 54.3							
Change in proficiency 2002–05: 7.4							
Piecewise model							
Schools with 2002 level <65							
Intercept	15.0	0.9	17.6	64.6	17.7	0.469	428
Level 02	–0.233	0.018	–12.6				
Schools with 2002 level 65–80							
Intercept	18.3	8.3	2.2	64.9	294.1		137
Level 02	–0.283	0.114	–2.5				
Schools with 2002 level >80							
Intercept	–3.6	11.8	–0.3	–1092.7	12.2		60
Level 02	–0.003	0.137	0.0				
Basic model							
Intercept	14.8	0.6	24.4	65.1	5.4	0.428	625
Level 02	–0.228	0.011	–21.6				

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Limited English proficiency							
Proficiency level 2002: 56.5							
Change in proficiency 2002–05: 6.30							
Piecewise model							
Schools with 2002 level <65							
Intercept	21.5	1.4	15.8	78.6	26.9	0.822	108
Level 02	−0.274	0.028	−9.7				
Schools with 2002 level 65–80							
Intercept	16.5	11.7	1.4	85.7	−149.1		37
Level 02	−0.193	0.162	−1.2				
Schools with 2002 level >80							
Intercept	24.0	16.4	1.5	81.4	−397.0		14
Level 02	−0.295	0.185	−1.6				
Basic model							
Intercept	20.9	0.9	22.4	80.9	8.0	0.630	159
Level 02	−0.258	0.016	−16.4				

TABLE D2
Middle schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level		Adjusted R^2	Number of observations
				Point estimate	Standard deviation		
All							
Proficiency level 2002: 69.0							
Change in proficiency 2002–05: 2.41							
Piecewise model							
Schools with 2002 level <65							
Intercept	12.7	1.4	9.1	79.5	45.7	0.642	101
Level 02	–0.160	0.025	–6.5				
Schools with 2002 level 65–80							
Intercept	13.6	3.2	4.3	85.1	122.4		140
Level 02	–0.159	0.044	–3.6				
Schools with 2002 level >80							
Intercept	7.0	5.8	1.2	95.0	–145.9		55
Level 02	–0.074	0.068	–1.1				
Basic model							
Intercept	10.3	0.7	14.8	90.1	12.8	0.308	296
Level 02	–0.114	0.010	–11.5				
White							
Proficiency level 2002: 77.3							
Change in proficiency 2002–05: 2.01							
Piecewise model							
Schools with 2002 level <65							
Intercept	29.8	3.0	9.9	69.3	31.5	0.598	27
Level 02	–0.430	0.051	–8.5				
Schools with 2002 level 65–80							
Intercept	11.8	3.4	3.5	93.2	249.7		145
Level 02	–0.126	0.046	–2.7				
Schools with 2002 level >80							
Intercept	7.2	4.0	1.8	97.4	–434.7		110
Level 02	–0.073	0.046	–1.6				
Basic model							
Intercept	13.2	1.1	12.6	91.1	14.4	0.289	282
Level 02	–0.145	0.014	–10.7				

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
African American							
Proficiency level 2002: 54.0							
Change in proficiency 2002–05: 3.69							
Piecewise model							
Schools with 2002 level <65							
Intercept	14.5	1.2	11.8	70.0	29.5	0.622	191
Level 02	–0.208	0.024	–8.5				
Schools with 2002 level 65–80							
Intercept	26.5	8.3	3.2	74.8	166.0		47
Level 02	–0.355	0.118	–3.0				
Schools with 2002 level >80							
Intercept	–68.3	50.9	–1.3	86.2	211.9		4
Level 02	0.792	0.596	1.3				
Basic model							
Intercept	12.9	0.9	13.9	75.5	11.7	0.299	242
Level 02	–0.171	0.017	–10.2				
Hispanic							
Proficiency level 2002: 56.5							
Change in proficiency 2002–05: 4.19							
Piecewise model							
Schools with 2002 level <65							
Intercept	12.5	2.3	5.5	87.4	144.4	0.748	39
Level 02	–0.143	0.045	–3.2				
Schools with 2002 level 65–80							
Intercept	25.1	11.8	2.1	76.3	–4,079.2		13
Level 02	–0.329	0.167	–2.0				
Schools with 2002 level >80							
Intercept	6.9	16.0	0.4	70.0	–46.1		3
Level 02	–0.098	0.181	–0.5				
Basic model							
Intercept	13.7	1.4	9.8	81.3	17.4	0.471	55
Level 02	–0.169	0.024	–7.0				

(continued)

TABLE D2 (CONTINUED)

Middle schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Economically disadvantaged							
Proficiency level 2002: 51.7							
Change in proficiency 2002–05: 3.71							
Piecewise model							
Schools with 2002 level <65							
Intercept	15.8	1.0	16.2	67.2	19.8	0.692	224
Level 02	–0.235	0.020	–12.0				
Schools with 2002 level 65–80							
Intercept	12.5	10.5	1.2	68.9	–133.1		36
Level 02	–0.181	0.151	–1.2				
Schools with 2002 level >80							
Intercept	–13.1	64.5	–0.2	124.8	14.7		2
Level 02	0.105	0.766	0.1				
Basic model							
Intercept	15.5	0.8	19.2	67.9	7.6	0.462	262
Level 02	–0.228	0.015	–15.0				

Note: There were not enough students with limited English proficiency or students with disabilities to run the piecewise regressions for them.

TABLE D3
High schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
All							
Proficiency level 2002: 80.1							
Change in proficiency 2002–05: 1.96							
Piecewise model							
Schools with 2002 level <65							
Intercept	13.7	3.7	3.8	137.2	–388.2	0.678	15
Level 02	–0.100	0.065	–1.5				
Schools with 2002 level 65–80							
Intercept	25.2	3.5	7.2	84.8	55.0		112
Level 02	–0.297	0.047	–6.3				
Schools with 2002 level >80							
Intercept	11.4	3.0	3.8	90.1	135.2		160
Level 02	–0.127	0.035	–3.6				
Basic model							
Intercept	19.6	1.0	19.7	89.0	9.0	0.526	287
Level 02	–0.221	0.012	–17.8				
White							
Proficiency level 2002: 84.7							
Change in proficiency 2002–05: 1.83							
Piecewise model							
Schools with 2002 level <65							
Intercept	14.6	5.0	2.9	107.5	–237.4	0.668	5
Level 02	–0.136	0.096	–1.4				
Schools with 2002 level 65–80							
Intercept	25.4	3.8	6.7	86.9	62.4		76
Level 02	–0.293	0.051	–5.8				
Schools with 2002 level >80							
Intercept	20.6	2.5	8.1	94.0	49.7		195
Level 02	–0.219	0.028	–7.8				
Basic model							
Intercept	17.9	1.0	18.1	94.3	10.3	0.493	276
Level 02	–0.190	0.012	–16.4				

(continued)

TABLE D3 (CONTINUED)

High schools

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
African American							
Proficiency level 2002: 70.5							
Change in proficiency 2002–05: 2.20							
Piecewise model							
Schools with 2002 level <65							
Intercept	20.0	3.0	6.8	79.5	68.1	0.589	68
Level 02	–0.251	0.053	–4.7				
Schools with 2002 level 65–80							
Intercept	20.3	5.6	3.6	78.1	135.1		79
Level 02	–0.260	0.078	–3.4				
Schools with 2002 level >80							
Intercept	11.3	8.8	1.3	76.3	–260.2		59
Level 02	–0.148	0.102	–1.5				
Basic model							
Intercept	19.5	1.2	16.1	79.4	9.7	0.506	206
Level 02	–0.246	0.017	–14.5				
Hispanic							
Proficiency level 2002: 72.0							
Change in proficiency 2002–05: 2.54							
Piecewise model							
Schools with 2002 level <65							
Intercept	17.4	4.3	4.0	99.1	586.7	0.763	17
Level 02	–0.176	0.077	–2.3				
Schools with 2002 level 65–80							
Intercept	19.5	9.6	2.0	81.8	–773.3		17
Level 02	–0.239	0.131	–1.8				
Schools with 2002 level >80							
Intercept	43.4	11.6	3.7	83.1	119.7		18
Level 02	–0.522	0.134	–3.9				
Basic model							
Intercept	24.0	1.9	12.6	80.5	12.4	0.718	52
Level 02	–0.298	0.026	–11.4				

Variable name	Coefficient	Standard error	*t*-statistic	Steady-state level: Point estimate	Steady-state level: Standard deviation	Adjusted R^2	Number of observations
Economically disadvantaged							
Proficiency level 2002: 69 .9							
Change in proficiency 2002–05: 1.73							
Piecewise model							
Schools with 2002 level <65							
Intercept	19.0	2.6	7.3	79.6	62.7	0.586	69
Level 02	−0.238	0.048	−5.0				
Schools with 2002 level 65–80							
Intercept	29.1	5.5	5.3	75.3	66.4		94
Level 02	−0.387	0.075	−5.2				
Schools with 2002 level >80							
Intercept	12.8	10.1	1.3	71.2	−317.3		48
Level 02	−0.179	0.116	−1.5				
Basic model							
Intercept	20.8	1.2	17.0	76.3	8.7	0.544	211
Level 02	−0.273	0.017	−15.9				

Note: There were not enough students with limited English proficiency or students with disabilities to run the piecewise regressions for them.

NOTES

The authors greatly appreciate the assistance from the Virginia Department of Education in providing the data used in the analysis and in answering questions about Virginia's performance monitoring system. They are particularly thankful for the support of Dr. Patricia I. Wright, Chief Deputy Superintendent of Public Instruction, and Bethann Canada, Director of Educational Information Management, Office of Technology and Human Resources. Valuable information came from a seminar discussing this report and related issues in Richmond on December 18, 2006. In addition, to Dr. Wright and Ms. Canada the participants were H. Douglas Cox, Assistant Superintendent, Division of Special Education and Student Services; Kathleen Smith, Director of the Office of School Improvement; Dr. Robert Triscari, Director of Assessment Development, Assessment and Reporting Division; and Susan M. Williams, Manager of Educational Applications, Education Information Systems. Dr. Donald Cymrot, Managing Director of Center for Education, CNA Corporation, provided valuable assistance in reviewing the project design, the technical analysis, several presentations, and the report. Elana Mintz provided very capable assistance in editing the first draft of the report, and Janet Thomason provided outstanding assistance in editing and handling technical issues in subsequent drafts.

1. For technical reasons, statistics are frequently reported for 70 percent of schools. Statistics that cover 70 percent of schools with values for a variable nearest the mean approximately describe the dispersion of individual values of one standard deviation above and below the mean. As the name implies, it is standard statistical practice to describe spreads of one standard deviation.

2. It is possible that such models have been used, but the results are unpublished and not widely known among state officials. The author hopes that this report will lead the states in the Appalachian Region to provide more information to the Regional Educational Laboratory about techniques used to make projections and that other regional educational laboratories will acquire similar information.

3. To limit the impact of "outliers," bins with four or fewer schools were combined with the next-highest bin if levels were below 50 percent and with the next-lowest bin if levels were above 50 percent. Also, some bins were empty. For example, no high schools had proficiency levels below 40 percent.

4. Later in the report the cross-sectional results discussed here are used to estimate time series relationships. Systematic year-to-year variation was so large, however, and the three-year change period so short, that the time series changes could not be used to definitively test the central hypothesis or estimate the change-level relationship.

5. Value-added measures have other useful features such as being able to accurately assess the performance of individual teachers and measure how well students progress *after* they enter a given school. A key reason that value-added measures have attracted so much attention is the intuitive appeal of holding a school accountable for students' performance only after they enter that school. This is especially relevant for limited English proficiency students who recently entered the United States. Models using either student-level or school-level data can hold student performance constant at the point that students enter a given school or estimate confidence intervals surrounding point estimates.

6. In the first draft of this report the same models were estimated using school data aggregated into the bins shown in figure 4, rather than using school-level data. The coefficients were similar to those shown here. The main difference is that, as expected going from 7 or fewer observations per regression to 250 or more, the standard errors of the estimates

and the adjusted R^2 were much smaller. The standard errors, and especially the adjusted R^2, would be further reduced by weighting the regressions by the square root of the number of students tested at a given school. Giving greater weight to schools with more test takers is appropriate because there is substantial variation in the number of test takers in different subgroups in different schools and the accuracy of the statistics for a given school increases in proportion to the square root of the number of students tested. However, this approach was not used because the coefficients produced by this weighting system also would reflect systematic differences in the rate of progress made in schools of different sizes, as discussed in appendix C.

7. The 0.005 standard error for the slope coefficient in equation 3 shows that when the change in proficiency falls by 1.89 percentage points (because the level increased by 10 percentage points), 95 percent of schools would exhibit a decrease in the change of between 1.87 and 1.91 percentage points ($2 \times 0.005 \times 10 = 0.02$; $[0.189 \times 10] - 0.02 = 1.87$, $[0.189 \times 10] + 0.02 = 1.91$). Applying the range of change around its mean to schools at the 60 percent proficiency level leads to the prediction that 95 percent of those schools would exhibit a change of between −0.4 and 16.8 percentage points ($1.6 \times 60/10 = 9.6$; $19.4 - 9.6 = 6.8$; $2.8 \times 60/10 = 16.8$; $16.4 - 16.8 = -0.4$). Using the confidence intervals for both the intercept and slope coefficients at the same time shows that 95 percent of schools at 60 percent proficiency in one year will have gains of between −1.8 and +14.2 percentage points the next.

8. These estimates also could be obtained from special tabulations of the data. However, the model produces these estimates automatically.

9. Equation C1 was selected after testing several models that differed in the number of linear segments and the range spanned by each segment. Equation C1 provided the best tradeoff between producing coefficients with small standard errors and producing results with the highest overall explanatory power (R^2).

10. The higher R^2 of the piecewise model suggests that the straight-line relationship differs at least slightly with different proficiency ranges. However, this is a very stringent specification test, and other tests indicate that the fit of the linear model is good.

11. Results for Hispanics and limited English proficiency students were not included because there were too few schools reporting data for these subgroups to produce meaningful results when broken down into the three groups. Similarly, results were omitted for proficiency groups if they contained fewer than 35 schools. This was the case for the white subgroup at proficiency levels below 65 percent in middle and high schools and for the middle school African American and disadvantaged students subgroups at proficiency levels above 80 percent.

12. Selecting which variables to include is not an exact science, as many variables are close substitutes. The method selected was designed to identify factors that are likely to have some effect on equation 1.

13. Using F-statistics would have produced technically superior estimates, but F-statistics are not as intuitively easy to understand as R^2 and would not have changed the conclusions in any way, as the F-statistic and R^2 are mathematically related.

14. A key change in the computation of growth for comparison against the safe harbor standard that would make a major difference in incentives would be to calculate the change in proficiency as proficiency averaged over the current year and the immediate past year divided by proficiency averaged over the

preceding two or three years. States could make this change themselves under the current No Child Left Behind Act. However, the standard—proficiency increases of 10 percent of the percentage not proficient—could still be too easy or too difficult to provide a meaningful indication of progress.

REFERENCES

Brookings Institution. (2002). *How well are American students learning?* Retrieved from http://www.brookings.edu/gs/brown/bc_report/2006/2006report.htm

Capps, R., Fix, M., Murray, J., Ost, J., Herwantoro, S., Zimmermann, W., et al. (2004, December). *Promise or peril: immigrants, LEP students, and the No Child Left Behind Act.* The Urban Institute. Retrieved from http://www.fcd-us.org/usr_doc/PromiseOrPeril.pdf

Center for Mental Health in Schools, University of California, Los Angeles. *Data on the plateau or leveling off effect of achievement test scores.* Information Resource. Retrieved from http://smhp.psych.ucla.edu/pdfdocs/plateau.pdf

Center for Education Policy. (2005). *Year 3 of the No Child Left Behind Act.* Retrieved from http://www.asu.edu/educ/epsl/EPRU/articles/EPRU-0504-120-OWI.pdf

Council of Chief State School Officers. (2006.) *Profiles of state accountability systems 2005–2006. Kentucky, Tennessee, Virginia and other states.* Retrieved from http://www.ccsso.org/projects/accountability_systems/State_Profiles/

Figlio, D., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review, 94*(3), 591–604.

Figlio, D., & Rouse, C. (2005). *Do accountability and voucher threats improve low-performing schools?* (NBER Working Paper No. 11597). Cambridge, MA: National Bureau of Economic Research.

Hamilton, L., McCaffrey, D., & Koretz, D. (2005). *Threats to the validity of measures of achievement gains.* Paper presented at the Maryland Assessment Research Center for Education Success Conference on Longitudinal Modeling of Student Achievement. Retrieved from http://www.marces.org/conference/Longitudinal/hamilton.ppt

Hanuchek, E. A., & Raymond, M. E. (2002, June). *Improving educational quality: How best to evaluate our schools?* Paper prepared for Education in the 21st Century: Meeting the Challenges of a Changing World, Federal Reserve Bank of Boston.

Jacobson, L., LaLonde, R., & Sullivan, D. (1993). Earnings losses of displaced workers. *American Economic Review, 83*(4), 685–709.

Jacobson, L., LaLonde, R., & Sullivan, D. (2004). The returns from community college schooling for displaced workers. *Journal of Econometrics, 125*(1–2), 271–304.

Kane, T. J., & Staiger, D. O. (2002a). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings Papers on Education Policy 2002.* Washington, DC: Brookings Institution.

Kane, T. J. & Staiger, D. O. (2002b). The promises and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives, 16*(4), 91–114.

Ladd, H. F. (2004, March). *Lessons from North Carolina's school-based accountability system.* Paper prepared for Education Finance Research Consortium. Retrieved from http://finance.tc-library.org/Content.asp?abstract=true&uid=1749

Lee, J. (2004, April). How feasible is adequate yearly progress (AYP)? Simulations of school AYP "Uniform Averaging" and "Safe Harbor" under the No Child Left Behind Act. *Education Policy Analysis Archives, 12*(14). Retrieved from http://epaa.asu.edu/epaa/v12n14/

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on gaps: An in-depth look into national and state reading and math outcomes.* Cambridge, MA: The Civil Rights Project at Harvard University.

Lewis, A. (2006, August). *Celebrating 20 years of research on educational assessment: Proceedings of the 2005 CRESST conference* (CSE Technical Report No. 698). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, University of California, Los Angeles.

Linn, R. (2000). Assessments and accountability. *Educational Researcher*, *29*(2), 4–16.

Linn, R. L. (2005, summer). *Fixing the NCLB accountability system* (CRESST Policy Brief No. 8). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, University of California, Los Angeles.

Linn, R. (2006, June). *Educational accountability systems* (CSE Technical Report No. 687). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Center for the Study of Evaluation, University of California, Los Angeles.

Loveless, T. (2005). Test-based accountability: the promise and the perils. In D. Ravitch (Ed.), *Brookings Papers on Education Policy 2002*. Washington, DC: Brookings Institution.

MassPartners for Public Schools. (2005). *Facing reality: What happens when good schools are labeled "failures"? Projecting Adequate Yearly Progress in Massachusetts schools.* Retrieved from http://www.massteacher.org/news/headlines/ayp_report.pdf

Moscovitch, E. (2004, March). *Projecting AYP in Connecticut schools.* Prepared for Connecticut Education Association. Retrieved from http://www.cea.org/nclb/upload/AYPCurtisFinal.pdf

No Child Left Behind Act of 2001. Title I–Improving the academic achievement of the disadvantaged. Public Law 107–110.

Novak, J. R., & Fuller, B. (2003, December). *Penalizing diverse schools? Similar test scores, but different students, bring federal sanctions* (Policy Brief No. 03–04)?" http://pace.berkeley.edu/reports/PB.03-4.pdf

Packer, J. (2004, July). *No Child Left Behind and Adequate Yearly Progress fundamental flaws: A forecast for failure.* Paper presented at the Center on Education Policy Forum on Ideas to Improve the Accountability Provisions Under the No Child Left Behind Act in Washington, DC.

Rose, L. (2002). *Hard facts regarding Adequate Yearly Progress (AYP) under the No Child Left Behind Act.* Special analysis for the Indiana Urban Superintendents Association, Indianapolis, IN.

Rothstein, R., Jacobsen, R., & Wilder, T. (2006, November). Proficiency for all—an oxymoron. In *Examining America's Commitment to Closing the Achievement Gap: NCLB and Its Alternatives.* Symposium conducted at Teachers College, Columbia University, New York.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*(4), 329–339.

Tennessee Department of Education. (2006, March). Proposal to the U.S. Department of Education: NCLB growth model pilot program. Nashville: Author.

Tennessee State Board of Education. (2005, January). *Master plan for Tennessee schools: Meeting the challenges of the 21st century* (FY2006). Nashville: Author.

United States Government Accountability Office. (2006, July). *No Child Left Behind Act: States face challenges measuring academic growth that education's initiatives may help addresss* (Report to Congressional Requesters No. GAO-06–661). Washington, DC: Government Printing Office.

Virginia Department of Education. (2000). *Virginia Standards of Learning Assessments* (Technical Report). Retrieved from http://www.pen.k12.va.us/VDOE/Assessment/VA01-Techrpt.pdf

Wiley, E. W., Mathis, W. J., & Garcia, D. R. (2005, September). *The impact of the Adequate Yearly Progress requirement of the federal "No Child Left Behind" Act on schools in the Great Lakes Region.* Education Policy Research Unit. Retrieved from http://epsl.asu.edu/epru/documents/EPSL-0509-109-EPRU.pdf

Wright, S. P., Sanders, W. L., & Rivers, J. C. (2005). Measurement of academic growth of individual students toward variable and meaningful standards. In R. W. Lissitz (Ed.), *Longitudinal and Value Added Modeling of Student Performance.* Maple Grove, MN: JAM Press.