## 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 | | Adjusted $R^2$ | Number of observations |
|---|---|---|---|---|---|---|---|
| | | | | Point estimate | Standard deviation | | |
| **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 | | Adjusted $R^2$ | Number of observations | Difference between math and reading steady-state level |
|---|---|---|---|---|---|---|---|---|
| | | | | Point estimate | Standard deviation | | | |
| 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

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.

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.

## 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: Point estimate | Steady-state level: Standard deviation | Adjusted $R^2$ | Number of observations |
|---|---|---|---|---|---|---|---|
| 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 | | Adjusted $R^2$ | Number of observations |
|---|---|---|---|---|---|---|---|
| | | | | Point estimate | Standard deviation | | |
| 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**

| | | | | Steady-state level | | | |
|---|---|---|---|---|---|---|---|
| Variable name | Coefficient | Standard error | *t*-statistic | Point estimate | 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.