Contemporary Economic Policy Advance Access published June 16, 2006

# HOW MUCH DO STUDENTS MATTER? APPLYING THE OAXACA DECOMPOSITION TO EXPLAIN DETERMINANTS OF ADEQUATE YEARLY PROGRESS

JOHN M. KRIEG and PAUL STORER*

*The federal government and many state governments have recently passed legislation that punishes school districts for not showing consistent improvement in standardized test scores. This article measures the extent to which school performance reflects student characteristics. After splitting schools in the state of Washington based on adequate yearly progress, the authors find that an overwhelming percentage of the difference between high- and low-performing schools is explained by characteristics beyond the control of school administrators. Thus legislation designed to penalize poorly performing schools may hurt students who are most in need of academic aid.* (*JEL* I2)

The No Child Left Behind Act (NCLBA) of 2001 is a reauthorization of the Elementary and Secondary Education Act (ESEA), the central federal law impacting precollegiate education. The ESEA, first enacted in 1965, implements Title I, the federal government's aid program for disadvantaged students. As the newest embodiment of the ESEA, the NCLBA has expanded the federal role in education and has become a focal point of education debate. Perhaps the most debated portion of the NCLBA is the mandatory annual testing of students in grades 3–8 that must have been implemented by the 2005–2006 academic year. These annual tests are used to determine if the district achieved adequate yearly progress (AYP). Districts failing to meet AYP for a number of consecutive years may receive sanctions ranging from loss of Title I funding through loss of local control of the district.

This empirical analysis is based on individual student results from a national test given in all Washington state schools during the 2001–2002 academic year. The authors use the method developed by Oaxaca (1973) to decompose the differences in average student test scores for those students in AYP schools versus non-AYP schools and determine the amount attributable to influences beyond a school's direct control. The authors' motivation is to determine the extent to which differences in student outcomes at Washington's AYP and non-AYP schools are explained by the characteristics of the students attending these schools rather than decisions made by teachers, principals, and administrators. Specifically, they determine how much of the performance of schools failing to make AYP can be attributed to the characteristics of the student body rather than to the quality of the school itself.

*This is a revision of a paper presented at the Western Economic Association 77th annual conference, Seattle, June 2002. The authors wish to thank Pete Bylsma and Lisa Ireland of the Washington State Office of the Superintendent of Public Instruction for providing the data used in this study through a data-sharing agreement. Helpful comments were received from two anonymous referees and by seminar participants at Simon Fraser University, and the 2002 PNREC and WEAI meetings. Storer acknowledges financial support from the College of Business and Economics at Western Washington University.

*Krieg:* Associate Professor, MS-9074, Dept. of Economics, Western Washington University, Bellingham, WA 98225-9074. Phone 1-360-650-7405, Fax 360-650-6315, E-mail John.Krieg@wwu.edu

*Storer:* Associate Professor, MS-9074, Dept. of Economics, Western Washington University, Bellingham, WA 98225-9074. Phone 1-360-650-6531, Fax 360-650-6315, E-mail Paul.Storer@wwu.edu

**ABBREVIATIONS**

AYP: Adequate Yearly Progress
CoreSS: Core Developmental Standard Score
ESEA: Elementary and Secondary Education Act
ITBS: Iowa Tests of Basic Skills
ITED: Iowa Tests of Educational Development
NCLBA: No Child Left Behind Act
WASL: Washington Assessment of Student Learning



Contemporary Economic Policy
(ISSN 1074-3529)

doi:10.1093/cep/byl003

This question is relevant for education policy and has recently received media attention. For instance, in 2002 Richard Rothstein of the *New York Times* wrote, "policy makers have been so anxious to punish school failure that they have worried too little about defining what failure is" (Rothstein 2002, p. B9). Clearly the remedial actions dictated by the NCLBA will be unsuccessful if the method of identifying poorly performing schools is flawed. Of most relevance for this work is the possibility that swapping all of the students at a non-AYP school with those at a school making AYP would simply result in the same test score outcomes for the students and reversal of the schools' AYP/non-AYP labels.

The results indicate that most of the poor performance of schools not making AYP is based on the characteristics of students and their families rather than on inputs typically controlled by school administrators. This suggests that sanctions based on test scores will inappropriately penalize schools because many identified as failing perform poorly not because of any institutional feature but because that school is populated by students without characteristics associated with success. Alternatively, it is possible that part of the explanatory power of student characteristics derives from a Tiebout (1956) effect in which some types of parents locate in areas served by high-quality schools. If true, this Tiebout sorting could bias the findings in favor of determining that school performance is based on the characteristics of students and their families. The authors conduct a test for the quantitative importance of this effect and discuss the implications of the findings for educational policy.

## I. THE NCLBA AND SCHOOL TESTING IN WASHINGTON

The NCLBA was signed into law by President George W. Bush on January 8, 2002, and its provisions will be phased in over a period of several years. The law places important conditions on the use of federal Title I funds targeted to aid students in high-poverty schools. States are required to assess the performance of schools and reward schools that perform well while prescribing corrective action for schools that fail to meet benchmarks set by law. No specific assessment instruments are prescribed, but these assessment methods must test performance of all public school students within a state in at least two core areas: reading/language arts and mathematics. The results of these tests must be stated in terms of proficiency levels of students rather than percentile scores.

The NCLBA requires school districts to bring all students to the "proficient" level in reading and mathematics by the 2013–14 school year. Individual schools must meet state AYP targets toward this goal for both their overall student population as well as for eight demographic subgroups.[1] Although the definition of AYP varies by state, Washington is one of a number of states that measure the percentage of a school's students who are proficient in any given year. Because the goal of the NCLBA is for all students to show proficiency in their subjects by the 2013–14 school year, the target percentage of proficient students in any school increases each year until it reaches 100% in 2013.

The state of Washington introduced the Washington Assessment of Student Learning (WASL), a statewide test of reading, writing, listening, and mathematics in 1997. The WASL is the state of Washington's diagnostic tool used to identify AYP. The WASL is an open-ended exam covering four distinct areas of learning: reading, writing, listening and mathematics and is given at the 4th-, 7th-, and 10th-grade levels. As will be described, failing or succeeding schools are determined based on the percentage of students in a school who meet the state-level measure of competency in WASL math and reading. In 2001 26.7% of 4th-graders, 19.5% of all 7th-graders, and 29.5% of all 10th-graders met all four WASL standards.[2] In that year, 9.8% of schools offering third grade and 28.6% of high schools did not make AYP.

For the school year examined in this article, AYP benchmarks in Washington were calculated by first determining the cumulative 12-year improvement needed between 2001–2002 and 2013–14 to have 100% proficiency by the end of this period. This total improvement was then evenly divided over the 12-year

1. The eight subgroups are: American Indian, Asian/Pacific Islanders, black, Hispanic, white, the disabled, limited English, and students from economically disadvantaged families.

2. See http://www.k12.wa.us/edprofile.

period. For example, in the 2001–2002 school year, 29.7% of fourth-grade students were rated as math proficient under the NCLBA. If this figure rises by 5.86 percentage points in each of the subsequent 12 school years, the goal of 100% proficiency would be attained by 2013–14. Thus, the benchmark for the 2002–2003 school year would be for 29.7% + 5.86% = 35.56% of a school's students to show proficiency on the WASL. A school with less than 35.56% of their overall student body demonstrating math proficiency in the 2002–2003 school year would be classified as not meeting AYP.[3] Furthermore, a school does not meet AYP if any of their congressionally identified demographic subgroups fail to achieve 35.56% proficiency.

The NCLBA prescribes specific penalties for schools receiving Title I funds that fail to meet AYP but allows states to determine the structure of penalties for non–Title I schools. For example, in the case of Title I schools that fail to make AYP two years in a row, students in the school must be allowed to transfer to schools in the same district that do make AYP. In this case, the NCLBA requires that up to 5% of the district's Title I funds must be used to pay for transfer students' transportation. Schools failing to show improvement over three years are required to provide supplemental educational services, including private tutoring. Those failing over a longer time period are required to take action that may include replacing teachers or administrators and, in the extreme, loss of local governance.

These provisions place student test performance in the forefront of federal education finance and policy. As a result, it is important to understand the implications of the NCLBA as it begins to come into full force in the 2005–2006 school year. To do this, the authors look at some of the determinants of student test performance to see how student composition affects the performance of a school on standardized tests. This is important because the provisions of the NCLBA are implicitly predicated on the assumption that differences in school resources, policies, and procedures are the key determinant of standardized test results.

3. Because the data used herein is from the 2001–2002 school year, this article reports the definition of AYP at that time. Since the 2001–2002 school year, Washington has changed the definition of AYP slightly. First, rather than having to achieve a linear progression towards 100% proficiency, AYP is now defined based on a three-year stair step where the baseline increase every three years.

## II. EXISTING EVIDENCE ON DETERMINANTS OF SCHOOLING OUTCOMES

An extensive body of literature has examined the effect of characteristics of schools and their students on individual educational outcomes. A review of the literature on educational inputs on outcomes is provided by Hanushek (1996). Hanushek notes that the Coleman Report (Coleman et al. 1966) was the beginning of a series of studies that concluded that "student achievement seems unrelated to standard measures of the resources going into schools." This growing body of literature that demonstrates little measurable effect of input measures, such as teacher salaries, student–teacher ratios, or total spending per student has been interpreted by some to mean that "schools don't matter."

Hoxby (2001) discusses the effect of family characteristics and notes that these factors are significantly related to student outcomes. Hoxby attributes a significant fraction of these effects to school choice proxy effects. For example, families with higher incomes have a wider range of neighborhood choices, and thus incomes are generally positively correlated with both school choice and educational attainment. Evidence on the importance of school choice effects has been mixed, however. Research by Cullen et al. (2005) attempts to control for nonrandom selection into voucher or charter school programs by comparing outcomes for winners and losers in Chicago Public Schools high school lotteries. The authors find no evidence of a positive impact of winning the lottery on student outcomes within the first few years of high school.

The focus of this article is somewhat different from the existing literature in that the authors seek to determine how much student characteristics matter for schools rather than how much school characteristics matter for individual students. The focus is consistent with the emphasis that the NCLBA places on accountability of schools and districts. The authors implement an empirical method that quantifies the proportion of the difference in average student outcomes at AYP and non-AYP schools that is attributable to factors beyond the direct control of schools. The authors acknowledge the point made by Hoxby that schools may have an indirect effect on the composition of their student body through school choice and sorting effects and attempt to control for this sorting as well.

## III. DESCRIPTION OF THE DATA AND EMPIRICAL METHODOLOGY

To understand the impact of student background and characteristics on the determination of a school's performance, this article examines student test performance on a battery of nationally administered tests. The Iowa Tests of Basic Skills (ITBS) are taken by third- and sixth-grade students, and the Iowa Tests of Educational Development (ITED) are given to students in the ninth grade. The Iowa tests are statewide, annual, standardized exams intended to identify a student's developmental level and measure annual academic growth. For grades before high school, these exams cover vocabulary, word analysis, reading comprehension, listening, language, mathematics, and social sciences. The high school exams cover specific phases of the same categories, such as mathematics concepts and mathematic computation.

Through a data-sharing agreement, the authors have been given access to individual ITBS/ITED test results and demographic information for all third-, sixth-, and ninth-grade students in Washington public schools.[4] These tests were given in the spring of 2001. According to the 2001 *Washington State Data Book* (Washington State Office of Financial Management 2001), the sample covers 99.3%, 98.8%, and 92.7% of the third-, sixth-, and ninth-grade student populations. The Iowa tests are scored in a variety of ways to make cross-individual comparisons possible. This article uses the core developmental standard score (CoreSS) which is an index of academic performance on all sections of the Iowa tests completed by the student.[5] For ease of comparison, CoreSS scores are normalized to mean zero and standard deviation of one.

Besides test scores, the ITBS/ITED files provide a wealth of information regarding each student. Race, migrant status, native language, the presence of a computer at home, and gender are gathered for each student. For sixth and ninth grades, the variables also include parental educational achievements, career/educational aspirations of students, and measures of student social behavior, such as alcohol use and sports participation. For third-graders, information regarding the amount of reading and television viewing are also available. Selected demographic data of observed students appear in Table 1.

In addition to individual student data, this article utilizes data on each school and district in the state. These data include aggregate demographic information at both the school and district level, the percent of students in a free or reduced lunch program in a school, as well as school and district graduation and dropout rates. The ITBS/ITED results measuring learning outcomes for students in a school are separate from the WASL results that the state uses to determine if that school makes or does not make AYP. The clear advantage of using the ITBS as the measure of student performance is that the authors avoid the effects of potential "teaching to the test" in WASL results. The WASL and ITBS/ITED tests are administered in consecutive grades (3, 6, and 9 for the ITBS/ITED and 4, 7, and 10 for the WASL) and cover similar material. A recent study by Joireman and Abbott (2001) compared the two tests and arrived at the following conclusions: "Students who score high on the ITBS Reading, Math, and Language Tests are likely to score high on the WASL Reading, Math and Writing tests. However, the sizes of these correlations are not so high as to conclude that the ITBS and WASL provide an identical assessment of student learning" (p. 8). This assessment confirms that the WASL and ITBS are related measures of student ability but the "teaching to the test" issue makes the ITBS a better choice for the present analysis.

The methodology is as follows. Using the WASL, the authors first determine whether schools did or did not make AYP. To measure the academic differences between AYP and non-AYP schools, the authors then employ

4. A large body of literature using ITBS/ITED results exists primarily in education journals. Publications using these exams in economics journals include Howell et al. (2002), who used the ITBS to examine school vouchers, and Farkas (1996), who used the ITBS to measure human and social capital levels.

5. A detailed description of the Iowa tests may be found online at www.uiowa.edu/~itp/itbs.htm. The authors used alternative measures of student performance based on the Iowa tests and found little qualitative difference than what is reported next.

## TABLE 1
Descriptive Statistics of Individual Students

| | 9th Grade | | | 6th Grade (4th-Grade Measure) | | | 6th Grade (7th-Grade Measure) | | | 3rd Grade | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Characteristic** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** |
| CoreSS | 0.227 (0.906) | > | 0.050 (0.944) | 0.259 (0.898) | > | −0.149 (0.895) | 0.222 (0.915) | > | −0.049 (0.966) | 0.187 (0.937) | > | −.416 (.895) |
| Male | 0.494 (0.499) | > | 0.483 (0.500) | 0.492 (0.500) | = | 0.473 (0.499) | 0.490 (0.500) | = | 0.486 (0.499) | 0.495 (0.499) | = | .489 (.499) |
| White | 0.824 (0.381) | > | 0.711 (0.453) | 0.788 (0.408) | > | 0.586 (0.493) | 0.813 (0.390) | > | 0.627 (0.484) | 0.757 (0.428) | > | .382 (.486) |
| Black | 0.026 (0.159) | < | 0.058 (0.233) | 0.038 (0.190) | < | 0.059 (0.236) | 0.033 (0.178) | < | 0.075 (0.264) | 0.053 (0.225) | < | .093 (.290) |
| Hispanic | 0.053 (0.225) | < | 0.100 (0.299) | 0.059 (0.237) | < | 0.242 (0.428) | 0.055 (0.227) | < | 0.155 (0.362) | 0.077 (0.265) | < | .379 (.485) |
| Computer at home | 0.918 (0.274) | > | 0.885 (0.319) | 0.900 (0.299) | > | 0.761 (0.427) | 0.901 (0.299) | > | 0.820 (0.384) | 0.876 (0.329) | > | .645 (.478) |
| Computer used for schoolwork | 0.818 (0.386) | > | 0.780 (0.414) | 0.716 (0.451) | > | 0.519 (0.500) | 0.728 (0.445) | > | 0.605 (0.489) | 0.353 (0.478) | > | .202 (.402) |
| Bilingual | 0.011 (0.105) | < | 0.033 (0.178) | 0.020 (0.140) | < | 0.055 (0.228) | 0.022 (0.148) | < | 0.059 (0.235) | 0.041 (0.199) | < | .169 (.375) |
| Gifted | 0.017 (0.128) | < | 0.028 (0.165) | 0.061 (0.239) | = | 0.067 (0.251) | 0.040 (0.195) | = | 0.040 (0.196) | 0.039 (0.194) | > | .024 (.152) |
| Migrant | 0.004 (0.064) | < | 0.007 (0.082) | 0.006 (0.078) | < | 0.041 (0.198) | 0.004 (0.062) | < | 0.021 (0.142) | 0.006 (0.076) | < | .071 (.258) |
| Serious alcohol problem | 0.183 (0.386) | = | 0.180 (0.385) | 0.052 (0.223) | < | 0.071 (0.257) | 0.075 (0.263) | = | 0.083 (0.276) | | | |
| Serious drug problem | 0.223 (0.416) | < | 0.235 (0.424) | 0.058 (0.234) | < | 0.079 (0.271) | 0.088 (0.283) | = | 0.098 (0.298) | | | |
| Goal: graduate from college | 0.399 (0.490) | > | 0.378 (0.485) | 0.538 (0.499) | > | 0.484 (0.500) | 0.536 (0.499) | > | 0.500 (0.500) | | | |
| Father graduated from college | 0.150 (0.357) | > | 0.125 (0.331) | | | | | | | | | |
| Mother graduated from college | 0.155 (0.362) | > | 0.124 (0.329) | | | | | | | | | |
| Plan on joining military after graduation | 0.050 (0.218) | = | 0.051 (0.220) | | | | | | | | | |
| Plan on working after graduation | 0.167 (0.373) | < | 0.178 (0.382) | | | | | | | | | |
| Often read for fun | | | | | | | | | | 0.669 (0.470) | > | .582 (.493) |
| Sometimes read for fun | | | | | | | | | | 0.152 (0.359) | = | .161 (.367) |
| Never watch TV | | | | | | | | | | 0.074 (0.261) | = | .077 (.267) |
| Watch TV less than 1 hour/day | | | | | | | | | | 0.188 (0.391) | = | .179 (.383) |
| Watch TV 5+ hour/day | | | | | | | | | | 0.188 (0.391) | < | .251 (.433) |
| N | 28,666 | | 30,392 | 20,330 | | 1,296 | 16,618 | | 14,703 | 33,783 | | 5,603 |

*Notes:* SDs in parentheses. > = < determined at the 95% level of significance.

the Oaxaca decomposition to individual students' ITBS/ITED CoreSS scores. Because the ITBS is not used in the determination of AYP, the authors avoid any mechanical simultaneity that might occur if one were to compare AYP and non-AYP schools using the same test employed by the state to determine AYP.

One challenge this approach presents is that sixth-grade students are found in two types of schools. Washington sixth-graders attend either middle schools that cover grades 6 through 8 or elementary schools that cover kindergarten through sixth grade. Because AYP is determined by the WASL test, which is administered in the fourth and seventh grades, sixth-graders in middle schools are determined to be in AYP schools based on the seventh-grade test, whereas sixth-graders in elementary schools have AYP determination based on the fourth-grade WASL test.

The advantage of using the Oaxaca decomposition is that it allows for a division of observed average tests scores between AYP and non-AYP schools into both an explained and an unexplained portion. To better understand this, suppose that the following linear relationship exists between the Iowa test score for student $i$, $T_i$, and some determinant of test performance, $X_i$ ($X_i$ might be family income, for example).

$$(1) \qquad T_i = \alpha + \beta X_i + \varepsilon_i.$$

Equation (1) allows for the possibility that AYP and non-AYP schools differ in their ability to translate student potential into test scores. Thus, different values of $\alpha$ and $\beta$ may exist for both groups of schools. If students are separated into two groups according to whether they are at successful or failing schools and the mean of the random error $\varepsilon_i$ is zero for both types of schools, the difference between average test scores at AYP and non-AYP schools can be written as:

$$(2) \quad \bar{T}_{AYP} - \bar{T}_{non\text{-}AYP} = \alpha_{AYP} - \alpha_{non\text{-}AYP} + \beta_{AYP}\bar{X}_{AYP} - \beta_{non\text{-}AYP}\bar{X}_{non\text{-}AYP}.$$

Adding and subtracting $\beta_{AYP}\bar{X}_{non\text{-}AYP}$ on both sides of equation (2) gives:

$$(3) \quad \bar{T}_{AYP} - \bar{T}_{non\text{-}AYP} = \alpha_{AYP} - \alpha_{non\text{-}AYP} + \bar{X}_{non\text{-}AYP}[\beta_{AYP} - \beta_{non\text{-}AYP}] + \beta_{AYP}[\bar{X}_{AYP} - \bar{X}_{non\text{-}AYP}].$$

Equation (3) was developed by Oaxaca to explain differences between groups and is implemented by using two separate least squares regressions to estimate the AYP and non-AYP parameters. A common application of this method is to test for evidence of discrimination based on race or gender in the determination of salaries. Naive analyses of discrimination often compare average incomes between groups, but this ignores the fact that average qualifications might also differ so that any observed differentials in average salaries could be entirely justified. The Oaxaca decomposition provides an objective means of disentangling the effects of true discrimination versus justified differences. In the case examined in this article, the authors also seek to improve on naive (and potentially misleading) comparisons of average test score performance at schools. To do this, the authors focus on the decomposition in (3) rather than the values of the individual coefficient estimates.

The three groupings of terms on the right-hand side of (3) represent three sources of potential differences in average test scores between AYP and non-AYP schools. The first term, $\alpha_{AYP} - \alpha_{non\text{-}AYP}$, is the difference in average performance of the two groups that is not related to any of the characteristics included in the regression. In a sense, the first term is a "pure" performance difference that could be caused by differences in school quality or in individual characteristics unmeasured by the researcher. A higher value of $\alpha_{AYP}$ may just mean that students do better at high-scoring schools regardless of personal or family characteristics.

The second term, $\bar{X}_{non\text{-}AYP}[\beta_{AYP} - \beta_{non\text{-}AYP}]$, captures differences in the way that measured student characteristics translate into test scores at the two types of schools. If AYP schools do a better job of converting the measure of input ability, $X$, into test performance, then $\beta_{AYP}$ would be greater than $\beta_{non\text{-}AYP}$. This second term computes differences in test scores that would arise between identical students attending an

AYP school and those attending a non-AYP school.

The final term, $\beta_{AYP}[\bar{X}_{AYP} - \bar{X}_{non\text{-}AYP}]$, indicates how variation in average student characteristics between schools affects mean test scores. An easy way to interpret this term is that it captures the differences in test scores that would result if all students who varied only in their measured characteristics attended the same quality of school. If this term is large in size, then most of the differences in test scores are explained by student characteristics only. The authors call the size of this factor the "explained" differences in test scores and focus on it to determine the propriety of measuring school success based on test scores.

In this article, the Oaxaca regressions linking student characteristics to test scores include many independent variables so the single-variable equations are modified accordingly. Also, the authors recognize that the choice to add and subtract $\beta_{AYP}\bar{X}_{non\text{-}AYP}$ from (2) is arbitrary and that they could have just as easily added and subtracted $\beta_{non\text{-}AYP}\bar{X}_{AYP}$. These two choices essentially correspond to two methods of decomposing the differences in student characteristics. The first technique (method 1) calculates the difference in average test scores that would occur if both schools had the coefficients of the AYP schools. This addresses the question of what would happen if the non-AYP students all transferred to the AYP schools and the coefficients of the "education production" process at the AYP schools was unchanged. Equivalently, the authors can imagine transferring the AYP students to non-AYP schools and seeing how much of a difference in average tests scores remains between the two groups. The authors refer to this technique as method 2. Given that there is no reason to prefer one approach over the other, the authors report the results of both methods.

One advantage of applying the Oaxaca method is that it decomposes differences in average test scores for the two groups into differences in the way AYP and non-AYP schools translate student ability into test performance versus variations in the average characteristics of students at AYP and non-AYP schools. For instance, consider the educational system as a production process in which students are the raw material provided to schools. Schools transform these raw materials into educated students through the use of the traditional inputs to the education production process (these inputs include teachers, training, program dollars, etc.). The Oaxaca technique allows for large differences in the characteristics of both the raw materials and the inputs at different schools and allows researchers to measure the importance of each.

The authors exclude from the regressions any measures of input quality (such as spending per pupil) that are directly under the control of the school administration. This means that the effects of these variables are manifested through their impacts on the estimated coefficients of the test score regressions for the two groups of schools. The authors are interested in seeing how much of the differences in average school performance is attributable to these coefficient differences versus differences in student characteristics. The larger the fraction explained by student characteristics the less it seems reasonable to either penalize failing schools for their performance or to reward high-scoring schools.

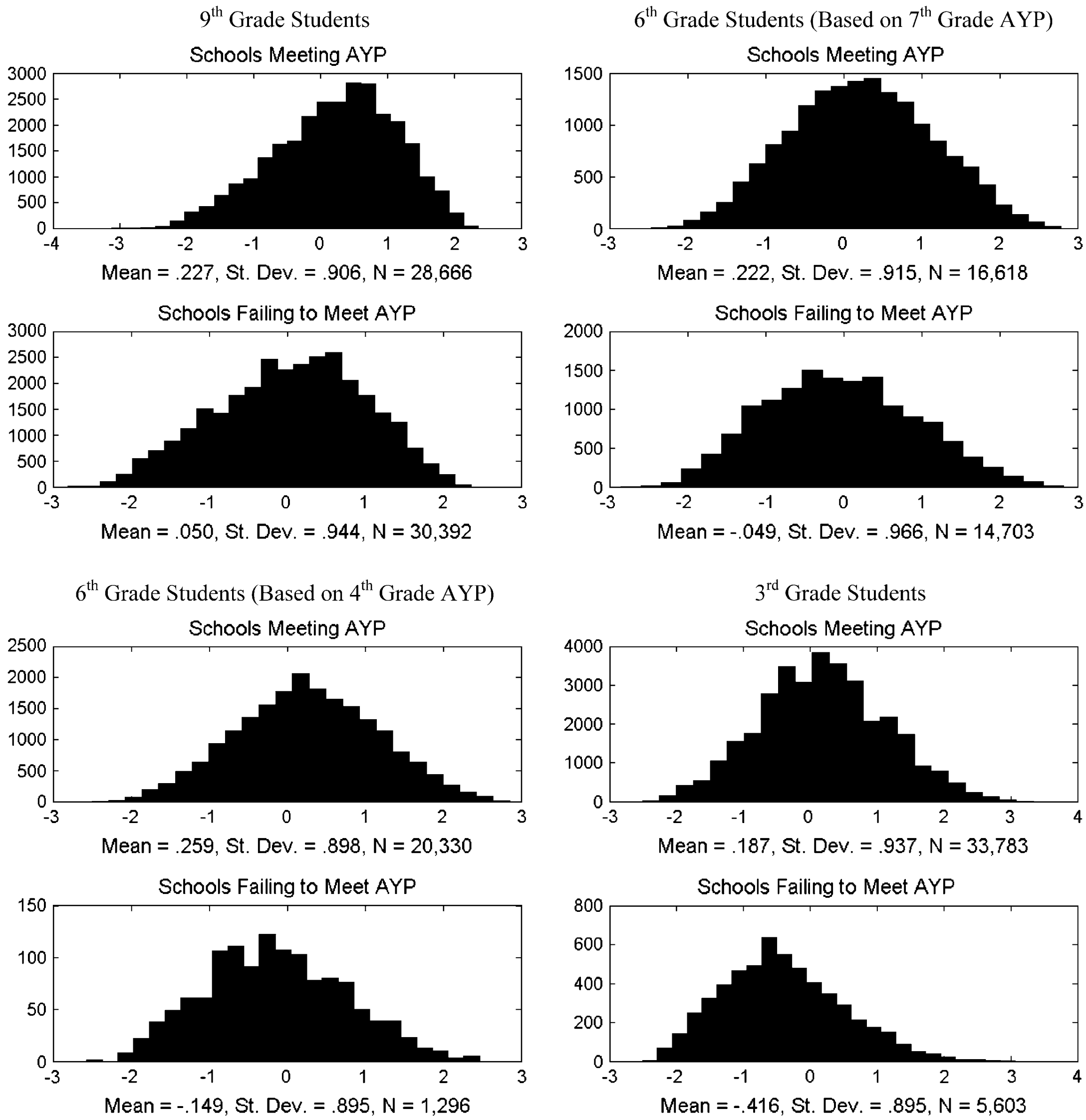
Of course, it is possible that high-scoring schools attract students from certain socioethnic groups because their parents shop around to find the best schools and purchase homes in the areas served by these schools. In this case, the differences in student composition (such as parental education levels) would not be entirely exogenous but rather partly determined by the quality of the school, and the Oaxaca decomposition would attribute too much of the difference in mean test scores to student characteristics. This type of sorting according to the characteristics of a local government jurisdiction was examined by Tiebout (1956) and, as a result, today bears his name. Hoxby (2001) discusses Tiebout sorting in the school choice context and argues that such choice by parents is one of the reasons why family characterics matter in student test score regressions. After discussing the main empirical results of the article, the authors conduct several tests to determine whether the results are sensitive to this type of self-selection effect.

## IV. EMPIRICAL RESULTS

### *A. Differences in Characteristics of Students, Schools, and Districts*

Figure 1 displays histograms for the distributions of individual ITBS/ITED CoreSS

**FIGURE 1**
Normalized CoreSS Histogram




test scores at the third, sixth, and ninth grade. In all cases, students at schools making AYP average statistically significantly higher scores than students at non-AYP schools. At the ninth-grade level, students in AYP schools average 0.177 standard deviations above students at non-AYP schools. This difference is substantially larger in the earlier grades. It is worth noting that some individual students at AYP schools are at the bottom of the CoreSS range, whereas there are also students at non-AYP schools who are at the top of the distribution.

Table 1 presents average student characteristics grouped by their school's AYP designation. Table 1 reveals that a much higher fraction of white students attend AYP schools and a higher fraction of black and Hispanic students attend non-AYP schools. Children at AYP schools also tend to have parents who are better educated, to have greater access to home computers, and to have a lower propensity to be from a migrant family or a family that does not speak English as their native language. AYP schools also tend to have fewer students with drug and alcohol problems at the lower grade levels, more students categorized as gifted, and more students who aspire to graduate from college. Although no direct measure of family prosperity is available, the home computer access variable is likely a proxy of household income as much as a measure of the actual direct contribution of computers to student performance. AYP schools have a greater proportion of students with home access to a computer, suggesting that more prosperous families attend these schools. As seen in Table 1, most of these differences are consistent across grades.

Table 2 provides similar information about the characteristics of AYP and non-AYP schools. In addition to the racial trends noted at the individual level, the percentage of students in the school that receive free lunches is much higher at non-AYP schools than at AYP schools. Average school enrollment is also higher at non-AYP schools. With the possible exception of school size, these characteristics are not chosen by the schools or districts and are thus not likely a measure of differences in local educational policies that affect school quality.[6]

Though it is clear that students in schools with higher WASL proficiency levels also tend to do well on the Iowa tests, the ultimate reason for this greater success is not yet apparent. Success on both measures may reflect factors beyond the schools' control (such as better-educated parents who are more able to help with homework or are more committed to the education process). This leads the authors to ask the question how much of the difference in average test scores found for these two groups can actually be explained by factors that school administrators can control and how much is explained by characteristics schools have little or no direct influence over. The authors now turn to the Oaxaca decomposition to quantify the answer to this question.

### *B. The Oaxaca Decomposition*

To conduct Oaxaca decompositions, the authors first estimate multiple-regression versions of equation (1) using separate ordinary least squares regressions for students attending the AYP schools and for students at non-AYP schools. Each grade level is examined separately and the students' CoreSS scores are used as the dependent variable. Although the authors are not interested in the individual regression coefficients themselves, the results are provided in the appendix.

The authors first conduct the analysis using regressions with only student and household characteristics as independent variables and then augment these regressions by including district and school demographics that are unlikely to be directly impacted by school policies (the percent of students receiving free or reduced lunches is an example of such a characteristic). Students are excluded from the analysis if they had missing values for any of the explanatory variables.[7] As already explained, there are two ways to conduct the Oaxaca decomposition, and the authors include results for both methods to check for robustness.

The results of the Oaxaca decomposition analysis are presented in Table 3. The first column of figures in the table provides the difference in average test scores at each grade level for students at AYP versus non-AYP schools. The remaining four columns provide the results of the Oaxaca decompositions. The first two columns of decomposition results are for regressions that only include individual student characteristics, whereas the second pair adds demographic characteristics of the school.

For example, the bottom row of Table 3 reveals that the average ninth-grade student in an AYP school scores 0.177 standard deviations on the ITBS/ITED above the average student in a non-AYP school. Of this 0.177 gap, differences in individual student characteristics explain between 0.120 and 0.122 (67.8–68.9%) of the difference depending on

6. Because school size may be determined by school administrators, the authors exclude school size measures from the Oaxaca decompositions.

7. Similar results were obtained when all observations were included and dummy variables assigned to cases with missing observations.

**TABLE 2**
School Characteristics

| | 9th Grade | | | 6th Grade (7th-Grade Meausre) | | | 6th Grade (4th-Grade Measure) | | | 3rd Grade | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Characteristic** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** | **Schools Making AYP** | | **Schools Failing to Make AYP** |
| Total school enrollment | 348.85 (404.98) | < | 1014.79 (599.68) | 495.64 (242.90) | < | 695.15 (228.78) | 462.27 (147.85) | = | 467.2 (132.20) | 392.05 (186.27) | < | 474.54 (151.59) |
| % of Asian students in school | 3.75 (5.48) | < | 7.85 (8.53) | 4.61 (5.78) | < | 7.18 (9.33) | 6.90 (6.40) | = | 8.49 (13.83) | 6.38 (7.36) | < | 8.37 (12.45) |
| % of black students in school | 4.62 (10.07) | = | 5.82 (8.98) | 2.51 (5.02) | < | 7.83 (12.30) | 4.32 (4.81) | = | 5.28 (6.73) | 5.04 (8.04) | < | 11.33 (17.46) |
| % of Hispanic students in school | 7.07 (11.88) | < | 12.28 (18.52) | 6.08 (8.53) | < | 19.85 (21.85) | 6.89 (9.84) | < | 32.64 (28.64) | 8.14 (11.22) | < | 32.55 (28.29) |
| % of white students in school | 79.18 (21.69) | > | 68.92 (23.60) | 84.13 (13.28) | > | 60.25 (25.04) | 79.43 (14.69) | < | 51.65 (25.80) | 77.42 (18.18) | > | 40.62 (26.66) |
| % of free lunch in school | 18.32 (22.30) | < | 27.28 (20.20) | 27.25 (18.67) | < | 44.89 (23.29) | 32.99 (20.32) | < | 63.14 (16.80) | 31.51 (23.19) | < | 63.34 (25.06) |
| # of schools | 435 | | 174 | 156 | | 93 | 402 | | 30 | 1094 | | 119 |

*Note:* > = < determined by a one-tailed, 95% level of significance.

**TABLE 3**
Percent and Size of CoreSS Explained between AYP and non-AYP Schools

| | | CoreSS Explained w/ Individual Characteristics | | CoreSS Explained w/ Individual and School Characteristics | | |
|---|---|---|---|---|---|---|
| | Difference in Mean CoreSS | Method 1 | Method 2 | Method 1 | Method 2 | *N* |
| 3rd-graders | 0.603 | 0.391 (64.8%) | 0.357 (59.2%) | 0.559 (92.8%) | 0.431 (71.5%) | 39,386 |
| 6th-graders (4th-grade test) | 0.407 | 0.293 (72.0%) | 0.242 (59.5%) | 0.392 (96.3%) | 0.255 (62.7%) | 21,626 |
| 6th-graders (7th-grade test) | 0.271 | 0.189 (69.7%) | 0.179 (66.1%) | 0.246 (90.8%) | 0.230 (84.9%) | 31,321 |
| 9th-graders | 0.177 | 0.120 (67.8%) | 0.122 (68.9%) | 0.127 (71.2%) | 0.120 (67.8%) | 59,058 |

*Note:* Method 1 subtracts $\beta_{AYP}\bar{X}_{non\text{-}AYP}$ from both sides of equation 2. Method 2 subtracts $\beta_{non\text{-}AYP}\bar{X}_{AYP}$ from both sides of equation 2.

the decomposition used. If district and school data are included, the range is 0.120–0.127 (67.8–71.2%).

A similar, pattern emerges when investigating the third- and sixth-grade data. Around 60% of the third-grade test score difference and between 60% and 70% of the sixth-grade difference is explained solely by differences in the characteristics of students and families at AYP versus non-AYP schools. When school and district variables are included, these figures range between 72% and 93% for the third grade and 63% to 96% for the sixth grade.

It should be stressed that the school and district variables that are included do not measure variables that teachers, principals, or district administrators are likely to influence but instead are composed solely of local demographic characteristics. For instance, the authors did not include school or district information on faculty-student ratios or instructional budgets per student. This allows the effects of policies under school administrator control to show up in coefficients that differ by school type (what the authors call the unexplained portion of the Oaxaca decomposition). Thus, the ninth-grade row in Table 3 indicates that between 67.8% and 71.2% of the test difference between AYP and non-AYP schools is explained by factors outside the direct control of administrators.

### C. *Sensitivity Tests*

One criticism of AYP is that poorly performing schools are required to make larger annual gains in the percentage of students achieving proficiency because all schools are required to reach 100% proficiency. Taking the example of math proficiency discussed earlier, a school that had the state average level of proficiency the previous year and added just 4 percentage points more students the following year (rather than the required 5.9 percentage points) would not make AYP, whereas a school that started well above the required percentage of students showing proficiency but remained at this same level the next year would remain an AYP school. Given these concerns about the determination of AYP, an alternative Oaxaca decomposition is estimated. Instead of the method actually used to determine AYP, this alternative measure of school performance divides schools into "high" performance and "low" performance schools based on the percent of students showing proficiency on the WASL. Specifically, for each grade, schools are ranked in order of the percent of students passing the WASL. Schools in the top quartile of these rankings are considered high-performing schools and those in the bottom quartile are considered low-performing. The decomposition then explains the ITBS/ITED differences between these sets of schools.

Table 4 displays the differences in mean test scores explainable by student characteristics in high- and low-performing schools. Although dividing schools using this method accentuates the score differences between high and low performing schools, Table 4 demonstrates similar findings to those found using the AYP split. Specifically, between 46.9% and 76.3% of test differences between high- and low-performing schools are accounted for solely by student characteristics. When

**TABLE 4**
Percent and Size of CoreSS Explained between High-Scoring and Low-Scoring Schools

| | Difference in Mean CoreSS | CoreSS Explained w/ Individual Characteristics | | CoreSS Explained w/ Individual and School Characteristics | | N |
|---|---|---|---|---|---|---|
| | | Method 1 | Method 2 | Method 1 | Method 2 | |
| 3rd-graders | 0.833 | 0.496 (59.5%) | 0.391 (46.9%) | 0.793 (95.2%) | 0.466 (55.9%) | 20,964 |
| 6th-graders (4th-grade test) | 0.626 | 0.412 (65.8%) | 0.303 (48.4%) | 0.563 (89.9%) | 0.402 (64.2%) | 10,993 |
| 6th-graders (7th-grade test) | 0.807 | 0.611 (75.7%) | 0.479 (59.4%) | 0.704 (87.2%) | 0.534 (66.2%) | 15,163 |
| 9th-graders | 0.654 | 0.499 (76.3%) | 0.455 (69.6%) | 0.627 (95.9%) | 0.489 (74.8%) | 23,377 |

*Note:* Method 1 subtracts $\beta_{AYP}\bar{X}_{non\text{-}AYP}$ from both sides of equation 2. Method 2 subtracts $\beta_{non\text{-}AYP}\bar{X}_{AYP}$ from both sides of equation 2.

school and district characteristics are added these fractions range from 55.9% to 95.9%. Hence, it appears likely that a majority of school success depends on student characteristics rather than school inputs.

Another sensitivity analysis can be conducted by comparing the columns for methods 1 and 2 in Tables 3 and 4. There is some evidence of lower effects of student characteristics using the method 2 decomposition, but this is not always the case. These differences are caused by different values of coefficient estimates for the two groups and the resulting changes in the conversion of average characteristic gaps into score gaps. However, it is likely that the true values are between those estimated by the two methods, and even in the most pessimistic scenarios, the authors still find that differences in average characteristics of students account for roughly half of AYP versus non-AYP differences in average student performance on the Iowa test scores.

### D. *Student Sorting*

The Tiebout (1956) hypothesis states that individuals will choose to live in localities that offer the amenities they seek. It is certainly true that parents take into account the quality of local schools when making residency decisions over cities within the country, school districts within a city, and individual school boundaries within a school district. This has the potential to impact observed links between test score performance and observable characteristics of students. Consider the case where parents with higher educational levels seek schools that perform well. In this case, a regression of student test scores on observable student and parent characteristics will overstate the impact of parental education on test scores because the parental education variables will capture both their own direct effect on student performance and the unobserved quality of the school correlated with parental education. This implies that the present results could overstate the extent to which school performance depends on student versus school characteristics.

To detect the influence of Tiebout sorting among schools *within* a district, the authors examine results for students in school districts without any alternative public school choice. They are unable to conduct similar tests for Tiebout sorting *between* districts (or cities) but note that Hoxby (2000, p. 1237) concluded that "much of the sorting of students by racial and income groups is at the school level, not the district." Hence, the authors can examine the robustness of the results to the type of student sorting that seems to be the most important.

Table 5 repeats the results for Table 3 using a restricted sample of schools. Specifically, this sample includes only schools that have no other public schools within their district that offer the grade being analyzed. In Washington, approximately one-third of ninth-grade students reside in a district with only one public ninth-grade option. Because middle and elementary schools are more common, there are fewer students without third- and sixth-grade choice within their public school districts. It is likely that districts with no choice at the elementary levels are fairly isolated and thus also benefit from a low level of

**TABLE 5**
Effect of Public School Choice in Own District AYP versus non-AYP Schools

| | Difference in CoreSS | CoreSS Explained w/ Individual Characteristics | | CoreSS Explained w/ Individual and School Characteristics | | *N* |
|---|---|---|---|---|---|---|
| | | Method 1 | Method 2 | Method 1 | Method 2 | |
| 3rd-graders | 0.626 | 0.403 (64.4%) | 0.471 (75.2%) | 0.413 (66.0%) | 0.970 (154.9%) | 4,579 |
| 6th-graders (4th-grade test) | 0.063 | 0.054 (85.7%) | 0.107 (169.8%) | 0.062 (98.2%) | 0.125 (198.4%) | 2,001 |
| 6th-graders (7th-grade test) | 0.311 | 0.194 (62.3%) | 0.211 (67.8%) | 0.231 (74.2%) | 0.376 (120.9%) | 7,574 |
| 9th-graders | 0.253 | 0.167 (66.0%) | 0.169 (66.8%) | 0.157 (62.1%) | 0.146 (57.7%) | 18,866 |

between-district choice, although the authors lack measures of school concentration and commute times that would allow them to make these conjectures more solid.[8]

Using this restricted sample, individual student and family characteristics account for roughly 66% of test score differences between AYP and non-AYP schools at the ninth-grade level. This compares to 67.8% for the entire sample of students. Based on only individual characteristics, this suggests that within-district Tiebout choice does not impact the results at the ninth-grade level. When adding district and school variables, the percent of the AYP/non-AYP CoreSS gap explained by student characteristics is 62.1% in districts without public school choice versus 71.2% in the full sample. Although this is a significant decrease in the restricted sample, the authors continue to find that individual and family characteristics are the most important factor in explaining average school test scores. The results for the third- and sixth-grade students without choice are less clear, and the authors attribute this to the more significant drops in sample sizes and possible rural/urban effects.

A second type of sorting may occur between private and public schools, and the authors address this in Table 6, which repeats the exercise of Table 5 but for districts in which there is no *private* school choice in the grade levels examined. Once again, the analysis focuses on ninth-grade students, but in this case, there are some more significant drops in the percentage of the score gap that is explained by student characteristics. This may be indicative of stronger sorting effects between public and private schools and may also indicate that factors like income matter more when a private school choice is available. Despite this, student characteristics account for at least 56% of the ninth-grade test score gap even in this restricted sample. These results suggest that sorting effects explain some (but far from all) of the effect of student characteristics on school performance. Future research should continue to focus on quantifying this effect.[9]

## V. CONCLUSIONS AND POLICY IMPLICATIONS

The results presented suggest that differences in average ITBS/ITED test scores between Washington state schools that pass or fail according to NCLBA AYP criteria largely reflect the characteristics of the students in these schools and not school policy choices. This raises doubts about both the efficacy

8. The authors are grateful to an anonymous referee for directing them to think about the effects of rural versus urban school locations on the degree of between-district choice.

9. Another type of sorting that may occur is parental choice between rural and urban districts. Using the National Center for Education Statistics definitions of rural school districts and defining an "urban" district as one that is within a Metropolitan Core Based Statistical Area, the authors also split the sample by schools making or failing to make AYP in both urban and rural areas. The results of this split were similar to those found in Tables 5 and 6. Specifically, the percent of CoreSS explained by Individual Characteristics in rural districts for third- through ninth-graders are 62.7%, 67.9%, 68.5%, and 62.0%. The corresponding percentages for urban students are 69.2%, 75.3%, 80.3%, and 75.2%. The similarities between these results and those presented in Tables 5 and 6 reinforce the estimates of the size of the sorting effect. The consistency of these results is probably not coincidental given that school districts with no public or private alternatives are likely to be in rural areas.

**TABLE 6**
Effects of Private School Choice in Own District AYP versus non-AYP Schools

| | | CoreSS Explained w/ Individual Characteristics | | CoreSS Explained w/ Individual and School Characteristics | | |
|---|---|---|---|---|---|---|
| | Difference in CoreSS | Method 1 | Method 2 | Method 1 | Method 2 | *N* |
| 3rd-graders | 0.534 | 0.295 (55.2%) | 0.368 (68.9%) | 0.363 (68.0%) | 0.515 (96.4%) | 7,599 |
| 6th-graders (4th-grade test) | 0.391 | 0.226 (57.8%) | 0.199 (50.9%) | 0.233 (59.6%) | 0.133 (34.0%) | 3,720 |
| 6th-graders (7th-grade test) | 0.339 | 0.301 (88.8%) | 0.298 (86.4%) | 0. 301 (88.8%) | 0.543 (160.1%) | 6,392 |
| 9th-graders | 0.126 | 0.079 (62.7%) | 0.071 (56.3%) | 0.071 (56.3%) | 0.077 (61.1%) | 21,810 |

and the fairness of penalizing schools that do not perform well on standardized tests. This could be used as a justification for basing rewards and sanctions on changes in test scores, rather than their levels, as is the case in the NCLBA. However, the very fact that differences in test scores may be due to factors beyond the control of school policies raises doubts about the link between the incentives in the NCLBA and student performance. Moreover, the results presented here suggest that changes in school test scores from year to year might reflect variation in the socioeconomic makeup of the school rather than the quality of the school.

The NCLBA partially addresses this concern through its requirement that a school must show evidence of improvement for all subgroups. These subgroups will include economically disadvantaged students and each of the major racial and ethnic groups. Trends for these groups are less likely to be affected by composition effects but the impact of composition effects is still not zero. One limit on this aspect of NCLBA is that results for subgroups are only examined when these groups exceed a certain fraction of the school population.

Another possible interpretation of the results is that the strong link between student performance and demographic characteristics may be due to Tiebout sorting among schools. In this case, the corrective measures introduced for poorly scoring schools may have several unintended effects. First, if many parents have already sorted themselves out of these schools, there may be relatively little demand to accept provisions to switch schools. Also, the NCLBA's provisions to leave failing schools may weaken the existing incentives to choose high-scoring schools or districts. If parents expect that they will be able to attend higher quality schools if their local school is found to be underperforming, they may decide not to pay the costs of living in a superior school area. This sets up a moral hazard problem in which the school transfer provisions of the NCBLA undermine the existing incentives to choose high-scoring schools.

**APPENDIX TABLE A.1**
Underlying Ordinary Least Squares Corresponding to Table 3

| | 3rd Grade | | 6th Grade (7th-Grade Measure) | | 6th Grade (4th-Grade Measure) | | 9th Grade | |
|---|---|---|---|---|---|---|---|---|
| | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** |
| Student Characteristics | | | | | | | | |
| Student is Male | .005 (.008) | −.019 (.020) | .002 (.011) | −.021* (.011) | −.006 (.010) | −.113*** (.041) | .054*** (.008) | .046*** (.008) |
| Student is Amer. Indian | −.221*** (.053) | −.006 (.101) | −.314*** (.065) | −.145** (.060) | −.101* (.053) | −.883** (.392) | .053 (.041) | −.070* (.041) |
| Student is Asian | .099** (.046) | .154 (.097) | .094* (.055) | .219*** (.054) | .233*** (.046) | −.610* (.369) | .029 (.036) | −.040 (.035) |
| Student is black | −.248*** (.047) | −.149 (.096) | −.310*** (.060) | −.246*** (.055) | −.171*** (.049) | −.994*** (.370) | −.218*** (.040) | −.350*** (.036) |
| Student is Hispanic | −.284*** (.046) | −.271*** (.088) | −.251*** (.057) | −.085 (.052) | −.103** (.047) | −.883** (.366) | .0007 (.037) | −.072** (.034) |
| Student is white | .030 (.043) | .182** (.088) | −.026 (.051) | .149*** (.049) | .115*** (.041) | −.611* (.362) | .143*** (.032) | .078** (.032) |
| Watch < 1 hr TV/day | .010 (.018) | .185*** (.044) | −.097*** (.026) | −.049 (.032) | −.086*** (.027) | .109 (.113) | −.054*** (.017) | −.023 (.019) |
| Watch 1 hr TV/day | .016 (.018) | .196*** (.044) | −.086*** (.026) | −.007 (.031) | −.098*** (.026) | .208* (.108) | −.029 (.018) | −.020 (.019) |
| Watch 2 hr TV/day | .067*** (.019) | .313*** (.044) | −.123*** (.025) | −.071** (.031) | −.123*** (.026) | .118 (.103) | −.087*** (.017) | −.045** (.018) |
| Watch 3 hr TV/day | .040** (.019) | .355*** (.045) | −.189*** (.026) | −.075*** (.031) | −.154*** (.026) | .199* (.103) | −.113*** (.018) | −.067*** (.019) |
| Watch 4 hr TV/day | .011 (.022) | .290*** (.051) | −.214*** (.029) | −.142*** (.033) | −.218*** (.028) | .078 (.108) | −.113*** (.021) | −.076*** (.021) |
| Watch 5+ hr TV/day | −.163*** (.018) | .181*** (.040) | −.340*** (.029) | −.217*** (.032) | −.315*** (.028) | −.014 (.106) | −.196*** (.021) | −.115*** (.021) |
| Computer at home | .234*** (.014) | .208*** (.023) | .096*** (.025) | .031 (.021) | .055** (.022) | .041 (.066) | .013 (.019) | .018 (.018) |
| Computer use at school | .096*** (.009) | −.005 (.026) | .185*** (.015) | .181*** (.015) | .158*** (.013) | .130*** (.049) | .135*** (.013) | .146*** (.012) |
| Student has email access | | | .107*** (.018) | .141*** (.018) | .116*** (.016) | .136** (.055) | .067*** (.013) | .033** (.013) |
| Held back 1+ grades | −.300*** (.015) | −.266*** (.032) | −.298*** (.024) | −.282*** (.023) | −.301*** (.023) | −.319*** (.076) | −.172*** (.015) | −.183*** (.014) |
| English sometimes spoken at home | −.047*** (.010) | −.041 (.026) | −.059*** (.015) | −.095*** (.015) | −.055*** (.012) | −.096* (.055) | .230*** (.020) | .176*** (.016) |
| English never spoken at home | −.364*** (.016) | −.193*** (.031) | −.246*** (.029) | −.207*** (.023) | −.214*** (.025) | −.101 (.075) | −.249*** (.021) | −.273*** (.016) |
| Parental help w/ homework monthly | .207*** (.015) | .241*** (.033) | .059*** (.021) | .103*** (.020) | .036** (.018) | −.099 (.067) | −.093*** (.009) | −.108*** (.009) |
| Parental help w/ homework weekly | −.009 (.013) | .096*** (.028) | −.177*** (.021) | −.133*** (.020) | −.230*** (.017) | −.270*** (.065) | −.296*** (.011) | −.298*** (.010) |
| Parental help w/ homework daily | −.278*** (.014) | −.134*** (.027) | −.458*** (.022) | −.377*** (.021) | −.478*** (.019) | −.465*** (.068) | −.486*** (.016) | −.488*** (.015) |
| Student is gifted | 1.090*** (.022) | 1.11*** (.065) | .850*** (.025) | 1.030*** (.025) | 1.008*** (.026) | 1.072*** (.104) | .411*** (.031) | .520*** (.023) |
| Student changed school this year | −.251*** (.014) | −.241*** (.029) | −.027** (.012) | −.005 (.013) | −.075*** (.017) | .038 (.072) | −.093* (.057) | −.131** (.058) |
| Student reads often for fun | .390*** (.014) | .307*** (.028) | | | | | | |

*continued*

**APPENDIX TABLE A.1 continued**

| | 3rd Grade | | 6th Grade (7th-Grade Measure) | | 6th Grade (4th-Grade Measure) | | 9th Grade | |
|---|---|---|---|---|---|---|---|---|
| | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP |
| Student reads sometimes for fun | .203*** (.016) | .113*** (.034) | | | | | | |
| Student reads little for fun | .000 (.021) | −.113 (.042) | | | | | | |
| Goal: high school graduate | | | .119 (.073) | .091 (.063) | .181** (.071) | −.036 (.249) | −.044 (.039) | −.051 (.038) |
| Goal: vocational school graduate | | | .156** (.079) | .110 (.069) | .219*** (.076) | −.095 (.270) | .142*** (.038) | .122*** (.036) |
| Goal: attend college | | | .186*** (.071) | .205*** (.061) | .259*** (.067) | .074 (.243) | .090** (.038) | .100*** (.037) |
| Goal: college graduate | | | .448*** (.069) | .427*** (.059) | .532*** (.067) | .318 (.239) | .292*** (.038) | .272*** (.037) |
| Goal: attend graduate program | | | .569*** (.070) | .489*** (.060) | .636*** (.067) | .343 (.240) | .381*** (.039) | .353*** (.037) |
| $<$ 1 hour homework per week | | | −.058 (.051) | −.029 (.048) | −.074* (.044) | .040 (.153) | .064** (.027) | .008 (.025) |
| 1 hour homework per week | | | −.044 (.051) | .019 (.047) | −.069 (.044) | .037 (.153) | .030 (.027) | .0008 (.025) |
| 2 hours homework per week | | | .034 (.051) | .059 (.048) | .017 (.044) | .159 (.154) | .085*** (.026) | .059** (.025) |
| 3 hours homework per week | | | .179*** (.051) | .163*** (.049) | .106** (.044) | .153 (.157) | .157*** (.026) | .119*** (.025) |
| 4−6 hours homework per week | | | .250*** (.051) | .278*** (.048) | .193*** (.044) | .299 (.156) | .208*** (.026) | .188*** (.025) |
| 7−9 hours homework per week | | | .325*** (.054) | .289*** (.052) | .256*** (.046) | .309 (.168) | .250*** (.028) | .227*** (.026) |
| 10+ hours homework per week | | | .338*** (.057) | .196*** (.056) | .128** (.051) | .094 (.184) | .184*** (.029) | .210*** (.028) |
| Serious absentee problem | | | −.123*** (.031) | −.061** (.026) | −.171*** (.031) | −.472*** (.108) | −.051** (.020) | −.068*** (.019) |
| Moderate absentee problem | | | .118*** (.016) | .150*** (.016) | .073*** (.015) | .055 (.057) | .033** (.013) | .051*** (.014) |
| Minor absentee problem | | | .073*** (.013) | −.014 (.019) | .079*** (.011) | .051 (.046) | .040*** (.012) | .069*** (.013) |
| Serious alcohol problem | | | −.226*** (.039) | −.239*** (.038) | −.173*** (.047) | −.028 (.151) | −.051** (.020) | −.068*** (.019) |
| Moderate alcohol problem | | | −.185*** (.030) | −.119*** (.030) | −.153*** (.040) | −.042 (.138) | −.015 (.016) | .025 (.016) |
| Minor alcohol problem | | | −.053*** (.018) | −.014 (.018) | −.053** (.022) | .059 (.078) | −.007 (.014) | .054 (.014) |
| Serious drug problem | | | .189 (.153) | .275** (.128) | −.127 (.155) | −.651 (.726) | .104*** (.019) | .071*** (.019) |
| Moderate drug problem | | | .301** (.151) | .325*** (.127) | −.099 (.155) | −.424 (.724) | .131*** (.016) | .118*** (.016) |
| Minor drug problem | | | .276* (.150) | .362*** (.125) | −.096 (.151) | −.383 (.717) | .122*** (.014) | .054*** (.014) |
| Serious cutting class problem | | | | | | | −.049** (.019) | −.015 (.018) |
| Moderate cutting class problem | | | | | | | .035*** (.014) | .075*** (.015) |
| Minor cutting class problem | | | | | | | .054*** (.012) | .088*** (.014) |
| Serious behavior problem | | | | | | | .085*** (.018) | .092*** (.017) |
| Moderate behavior problem | | | | | | | .120*** (.015) | .150*** (.015) |
| Minor behavior problem | | | | | | | .101*** (.014) | .113*** (.014) |
| Play 1 sport | | | | | | | −.026** (.012) | −.026** (.011) |
| Play 2 sports | | | | | | | −.050*** (.013) | −.024** (.012) |

*continued*

**APPENDIX TABLE A.1 continued**

| | 3rd Grade | | 6th Grade (7th-Grade Measure) | | 6th Grade (4th-Grade Measure) | | 9th Grade | |
|---|---|---|---|---|---|---|---|---|
| | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP | Schools Making AYP | Schools Not Making AYP |
| Play 3 sports | | | | | | | –.043*** (.014) | –.038*** (.014) |
| Play 4+ sports | | | | | | | –.051*** (.013) | –.037*** (.012) |
| GPA: A | | | | | | | .475*** (.055) | .697*** (.059) |
| GPA: A– | | | | | | | .102* (.054) | .365*** (.059) |
| GPA: B | | | | | | | –.112** (.054) | .166*** (.059) |
| GPA: B– | | | | | | | –.276*** (.054) | .004 (.059) |
| GPA: C | | | | | | | –.314*** (.055) | –.017 (.060) |
| GPA: C– | | | | | | | –.401*** (.055) | –.075 (.061) |
| GPA: D | | | | | | | –.370*** (.059) | –.127** (.064) |
| GPA: F | | | | | | | –.452*** (.064) | –.085 (.068) |
| Plan to work | | | | | | | –.196*** (.016) | –.114*** (.018) |
| Plan to attend college | | | | | | | –.125*** (.016) | –.055*** (.017) |
| Plan to attend voc. school | | | | | | | –.172*** (.023) | –.045* (.023) |
| Plan to enter military | | | | | | | –.125 (.023) | –.103*** (.023) |
| Plan home | | | | | | | –.156*** (.031) | –.039 (.030) |
| Mother: no high school | | | | | | | –.031 (.019) | –.015 (.017) |
| Mother: high school | | | | | | | .053*** (.015) | .068*** (.014) |
| Mother: vocational school | | | | | | | .080*** (.015) | .112*** (.015) |
| Mother: attend college | | | | | | | .077*** (.018) | .088*** (.019) |
| Mother: graduate college | | | | | | | .152*** (.016) | .186*** (.017) |
| Mother: graduate school | | | | | | | .138*** (.019) | .179*** (.020) |
| Father: no high school | | | | | | | –.091*** (.018) | –.063*** (.017) |
| Father: high school | | | | | | | –.036** (.014) | –.034** (.013) |
| Father: vocational school | | | | | | | .008 (.015) | .006 (.014) |
| Father: attend college | | | | | | | .020 (.018) | –.003 (.018) |
| Father: graduate college | | | | | | | .089*** (.015) | .129*** (.016) |
| Father: graduate school | | | | | | | .104*** (.017) | .141*** (.018) |
| District Characteristics | | | | | | | | |
| % district bilingual | .011*** (.001) | –.003** (.001) | –.003 (.002) | .007*** (.001) | .008*** (.001) | .004 (.006) | .0004 (.001) | .005*** (.001) |
| District dropout % | –.002*** (.0006) | .001 (.001) | –.002* (.001) | –.0008 (.0008) | –.009*** (.002) | –.011 (.007) | –.001 (.001) | –.002*** (.0005) |
| District graduation % | .0009*** (.0003) | –.002*** (.0008) | .0002 (.0003) | –.002*** (.0006) | –.0006 (.0006) | .004 (.003) | .0002 (.0004) | –.001*** (.0005) |

*continued*

**APPENDIX TABLE A.1 continued**

| | 3rd Grade | | 6th Grade (7th-Grade Measure) | | 6th Grade (4th-Grade Measure) | | 9th Grade | |
|---|---|---|---|---|---|---|---|---|
| | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** | **Schools Making AYP** | **Schools Not Making AYP** |
| School Characteristics, Cont. | | | | | | | | |
| % district migrant students | .0005 (.001) | −.003 (.002) | .005** (.002) | −.0006 (.002) | .001 (.002) | −.004 (.005) | .0001 (.001) | −.008*** (.001) |
| % district special education | −.018*** (.003) | .003 (.008) | −.023*** (.003) | .013** (.006) | −.029*** (.004) | −.006 (.014) | −.017*** (.002) | .015*** (.004) |
| School % Asian | .001 (.0006) | −.0001 (.001) | −.002 (.001) | .002 (.0009) | .003*** (.001) | .012*** (.003) | .006*** (.0009) | .004*** (.0007) |
| School % Indian | .003** (.001) | −.002** (.0009) | −.009*** (.002) | −.002*** (.0009) | .007*** (.002) | −.003 (.024) | −.004*** (.0009) | −.0006 (.0007) |
| School % black | .0007 (.0007) | .0002 (.001) | .0001 (.001) | −.003*** (.0009) | −.004** (.001) | .002 (.004) | −.002** (.001) | −.003*** (.0007) |
| School % Hispanic | −.005 (.0008) | .002 (.001) | .0006 (.002) | −.004*** (.001) | −.003** (.001) | .002 (.004) | −.002 (.001) | .001 (.0008) |
| School % male | −.001 (.002) | .017*** (.004) | −.002 (.002) | −.001 (.003) | .006*** (.002) | −.003 (.015) | −.005*** (.001) | .003 (.002) |
| School % free/reduced lunch | −.006*** (.0002) | −.002*** (.0005) | −.002*** (.0005) | −.002*** (.0003) | −.004*** (.0003) | −.0009 (.002) | −.003*** (.0003) | −.002*** (.0003) |
| $N$ | 33,783 | 5,603 | 16,618 | 14,703 | 20,330 | 1,296 | 28,660 | 30,392 |
| $R^2$ | .316 | .329 | .365 | .458 | .363 | .418 | .494 | .510 |

*Notes:* All regressions contain a constant. *** (**) [*] represent statistical significance at the 1% (5%) [10%] levels.

## REFERENCES

Coleman, J. S., et al., *Equality of Educational Opportunity.* Washington, DC: U.S. Department of Health, Education and Welfare, 1966.

Cullen, J., B. Jacob, and S. Levitt, "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools," *Journal of Public Economics*, 89(5–6), 729–60.

Farkas, G., *Human Capital or Cultural Capital? Ethnicity and Poverty Groups in an Urban School District.* New York: Aldine de Gruyter, 1996.

Hanushek, E., "School Resources and Student Performance," in *Does Money Matter?* edited by Gary Burtless. Washington, DC: Brookings Institution Press, 1996.

Howell, W. G., P. J. Wolf, D. E. Campbell, and P. E. Peterson, "School Vouchers and Academic Performance: Results from Three Randomized Field Trials," *Journal of Policy Analysis and Management*, 21(2), 2002, 191–217.

Hoxby, C., "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90(5), 2000, 1209–38.

———, "If Families Matter Most, Where Do Schools Come In?" in *A Primer on America's Schools*, edited by Terry M Moe. Stanford, CA: Hoover Institution Press, 2001.

Joireman, J., and M. L. Abbott, *The Relationships between the Iowa Test of Basic Skills and the Washington Assessment of Student Learning in the State of Washington.* Technical Report #2, Washington School Research Center, Lynnwood, 2001.

Oaxaca, R., "Male-Female Differentials in Urban Labor Markets," *International Economic Review*, 14(3), 1973, 693–709.

Rothstein, Richard, "Defining Failed Schools Is Harder Than It Sounds," *New York Times*, July 3, 2002.

Tiebout, C., "A Pure Theory of Local Public Expenditures," *Journal of Political Economy*, 64(5), 1956, 416–24.

Washington State Office of Financial Management, *2001 Washington State Data Book.* Olympia: State of Washington, 2001; available online at www.ofm.wa.gov/databook/index.htm.