ARE STUDENTS LEFT BEHIND? THE DISTRIBUTIONAL EFFECTS OF THE NO CHILD LEFT BEHIND ACT

John M. Krieg
Department of Economics
Western Washington University
MS 9074
Bellingham, WA 98225
john.krieg@wwu.edu

Abstract
The No Child Left Behind Act imposes sanctions on schools if the fraction of students demonstrating proficiency on a high-stakes test falls below a statewide pass rate. While the motivation behind this system is improved public school performance, it also provides incentives for schools to focus educational resources on the marginal student rather than those on the tails of the ability distribution. Using statewide, student-level panel data, students on the tails of the ability distribution, especially high-ability students, are demonstrated to score below expectations if their school is in danger from No Child Left Behind sanctions.

1. INTRODUCTION

Demands for school accountability and education reform culminated in the No Child Left Behind (NCLB) Act, the 2002 reauthorization and expansion of the Elementary and Secondary Education Act. NCLB blends two prevailing schools of thought on improving K–12 education. By holding school districts accountable for student performance and by providing expanded educational choices for students in failing schools, the proponents of NCLB hope to improve overall educational quality. However, the structure of NCLB also provides incentives to reduce academic achievement for some groups of students. This article describes these incentives and documents a trade-off lowering academic growth for students at the tails of the ability distribution in favor of those in the center.

The NCLB Act institutes a system of performance goals that, if not met, trigger sanctions increasing in severity as schools and districts fail to meet those goals. Yet, as Ladd (2001) has suggested, any performance-based system suffers from a number of potential pitfalls. For instance, important objectives not emphasized by NCLB are likely to receive less attention under a performance-based system.[1] Second, when goals are translated into empirical measures, there may be a weak connection between the goals and the measures. For example, the presence of high-stakes exams encourages teaching specifically to the content of the exams, thereby improving measured achievement without broader academic growth.[2]

In addition to these concerns, NCLB generates incentives for teaching to specific subsets of students, which creates the possibility that students make differential academic gains based upon their propensity to aid their school's ability to satisfy NCLB. The NCLB Act requires schools to make adequate yearly progress (AYP) toward educating all students. As mandated by NCLB, AYP in the state of Washington is determined annually for each school by measuring the percent of students demonstrating proficiency relative to a statewide performance standard on a high-stakes exam. Schools with too few students demonstrating proficiency—that is, schools with the percent of proficient students below the state's required pass rate—do not make AYP and are subject to sanctions under NCLB. By focusing on a binary pass/no pass outcome, NCLB provides incentives to direct educational resources toward students on the margin of demonstrating proficiency at the expense of students who either have little probability of demonstrating proficiency or are nearly assured of

1. A short list of these goals includes physical education, nutrition, social sciences, etc.
2. Jacob (2005) finds that gains made on high-stakes tests are not mirrored in low-stakes tests, and the gains that are made on high-stakes exams appeared to be due to improvements in test-specific skills. Klein et al. (2000) make a similar finding for eighth graders when comparing the Texas high-stakes test and the National Assessment of Educational Progress.

doing so. For instance, when deciding which extracurricular activities to provide, a school administrator, hoping to raise the fraction of students meeting the reading standard, may opt for a reading program that targets marginal students rather than a program for accelerated readers. Alternatively, an administrator may abandon curricula intended to help the lowest-tier students in favor of one that is more appropriate for students marginally below the standard. Administrators may assign marginal students to strong teachers and other students to weaker teachers. Whatever their specific response, this behavior, identified in this article as "strategic instruction," can lead to lower achievement for students on the tails of the ability distribution.

To test for strategic instruction, I sort students by their scores on a low-stakes exam and by their school's potential to experience sanctions under NCLB. I then compare high-stakes exam performance of high- and low-ability students at schools that either are or are not facing NCLB sanctions. The resulting evidence is consistent with the presence of strategic instruction. Specifically, students at either tail of the low-stakes exam, if they are at schools that are more likely to be sanctioned under NCLB, gain less on the high-stakes exam than what would have been expected if their school did not face sanctions. Further, students at the center of the low-stakes test distribution do better than expected on the high-stakes exam if they attend schools that are likely to be sanctioned. These effects increase as schools face more severe sanctions, and they do not appear to be related to unobservable characteristics of the student's school. The effects are robust to student attrition, trimming the sample to eliminate outliers, and different estimation techniques.

A handful of researchers have investigated strategic instruction. Chakrabarti (2007) uses disaggregated school-level data to analyze the behavior response of schools threatened under Florida's opportunity scholarship program. This program predates NCLB and offers similar incentives to school administrators. Under this program, schools failing for two out of four years must provide students with vouchers. Chakrabarti argues that the incentive this program creates is to focus on students who are marginally below the threshold required to pass Florida's high-stakes test. When compared with students at similar but nonthreatened schools, Chakrabarti finds that marginal students at threatened schools improve performance. Further, Chakrabarti argues that the entire test distribution moves to the right, with larger moves for marginal students. Chakrabarti's evidence suggests that high- and low-ability students appear not to suffer under Florida's program.

Burgess et al. (2005) examine school accountability for secondary students in the United Kingdom. If strategic instruction occurs, schools with a higher proportion of marginal students will have a greater incentive to divert resources from students at the tails of the ability distribution. Indeed, these authors find that as the proportion of marginal pupils increases, all students lose relative to the most able, but the lowest ability group loses the most. One possible explanation for the relative stability of the most able students is that UK schools have overlapping catchment zones, leading to school competition for the best students.

Using pre-NCLB Texas data on individual students, Reback (2008) finds that schools respond to the Texas accountability system with measures helping low-performing students and specific, targeted measures toward students that are critical to the school's accountability ratings. The data allow the author to compare students within buildings, and he finds that those gaining most academically are also those who have the highest probability of increasing their school's rankings. In contrast, relatively high-achieving students perform worse than expected if their performance is unlikely to affect their school's ratings. It should be noted that Reback's sample excludes up to one-half of the state's highest scoring students because these students would be unlikely to demonstrate increased year-to-year gains under the Texas testing system.

The current article offers a number of innovations over past research. First, it uses student-level data on a pretest that is not the object of NCLB evaluations. This pretest is used as an explanatory variable in regressions on student-level high-stakes test results. Student-level results on this pretest are known to every school administrator prior to the beginning of the AYP cycle and are a natural tool for administrators to base resource decisions. If strategic instruction does take place, one would expect it to be based on data gathered by administrators, such as this pretest. Further, the data used in this work cover a high percentage of all general education students in the state of Washington and control for sample selection bias potentially caused by those general education students who remain unobserved. Finally, this article observes students both before and after the enactment of NCLB, which allows for direct comparisons of the impacts of the law.

Before proceeding, it should be noted that strategic instruction may not be an inefficient outcome. If, prior to NCLB, schools overexpended resources on students at the tails of the ability distribution, then NCLB incentives may improve overall resource allocation. As suggested by Chakrabarti (2007), if building administrators respond to NCLB by introducing better targeted teaching techniques, better curriculum, or a more efficient use of resources, then NCLB may raise the level of all students.

2. NCLB AND STUDENT TESTING IN WASHINGTON

The NCLB Act requires school districts to bring all students to the proficient level in reading and mathematics by the 2013–14 school year. In the meantime, individual schools must meet state AYP targets directed toward this goal for their overall student population as well as for eight demographic subgroups: American Indian, Asian/Pacific Islanders, black, Hispanic, white, special education, limited English, and economically disadvantaged students. To make AYP, the state of Washington measures the percentage of a school's students in each of these nine groups who demonstrate proficiency on a high-stakes exam and compares this with the state-imposed required pass rate. For a school to make AYP, the percentage of the total student body, as well as the percentage of each subgroup, must be above the required pass rate. As designed, a single student can be a member of many groups and therefore affect a school's ability to make AYP multiple times. For instance, an Asian, limited-English student from an economically disadvantaged family would be represented in the overall student body as well as three of the eight demographic subgroups. If this student fails to demonstrate proficiency on the high-stakes test, then this failure is represented in the overall calculation as well as the calculation of the three subgroups.

As part of a move toward educational accountability, in 1997 the state of Washington introduced the Washington Assessment of Student Learning (WASL), a statewide test of reading, writing, listening, and mathematics.[3] The WASL is Washington's high-stakes diagnostic tool used to identify AYP. At the fourth, seventh, and tenth grades (the grades examined in this article), the WASL tests mathematics, reading, and writing. In order to avoid complications that arise when combining scores from tests of different subjects, this article analyzes only the WASL math results.[4] In order to make consistent cross-year comparisons, student-level WASL scores have been normalized to mean zero and variance one within each grade and year.[5]

The opportunities for strategic instruction to occur under NCLB are vast and might involve resource shifting both within and between subgroups. For instance, a school with a high percentage of Hispanics demonstrating proficiency but too few American Indians could reduce resources dedicated to Hispanic students in favor of American Indian students. Alternatively, this school may reallocate resources from those successful American Indians to those who are less successful. Regardless of which strategy is chosen, under strategic instruction it is clear that those students with a high probability of

3. Starting in 2004, the listening test was eliminated and a science test was introduced in some grades.
4. Indeed, the vast majority of schools failing to make AYP in 2004–5 did so because of a failure to achieve the required pass rate in mathematics. In 2005, of the 207 Washington buildings failing to make AYP, 161 were due to poor math scores. Those failing to make AYP for reasons other than math are classified as AYP schools for the purposes of analysis in this article.
5. The mathematics WASL test is scored by giving each test item a certain number of possible points. These points are then summed. In 2002, the first year in this sample, .253 percent of Washington students scored 353 points—the highest possible WASL score.

demonstrating proficiency have fewer resources available. Because of this, this article focuses on each school's overall student body rather than strategic instruction between subgroups.

Required pass rates in Washington are calculated by first determining the cumulative twelve-year improvement needed between 2001–2, when NCLB was implemented, and 2013–14, in order to have 100 percent of all students demonstrate proficiency at the end of this period. This total improvement is then evenly divided over the twelve-year period. For example, in 2001–2, 29.7 percent of fourth-grade students were rated as math proficient under NCLB. If this figure rises by 5.86 percentage points in each of the subsequent twelve school years, the goal of 100 percent proficiency would be attained by 2013–14. Thus the mathematics pass rate required to make AYP in the 2002–3 school year was 29.7 percent + 5.86 percent = 35.56 percent. A school with less than 35.56 percent of its overall student body (or of any subgroup) demonstrating math proficiency in 2002–3 would be classified as not meeting AYP.[6] Finally, AYP is granted only if 95 percent of all continuously enrolled students at each grade level take the WASL. In 2006, 45.1 percent of fourth graders, 38.6 percent of seventh graders, and 45.1 percent of tenth graders demonstrated proficiency in math, reading, and writing. In that year, 10.3 percent of schools offering fourth grade, 21.8 percent of schools offering seventh grade, and 35.9 percent of high schools had insufficient students demonstrating proficiency to be above the required pass rate and hence did not make AYP.

The NCLB Act prescribes specific penalties for schools receiving Title I funds failing to meet AYP, but it 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 for two years in a row, students in the school must be allowed to transfer to a school in the same district that makes AYP. In this case, NCLB requires up to 5 percent of the district's Title I funds 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 replace teachers or administrators, and, in extreme cases, face loss of local governance. This increased scope of sanctions for schools failing to make AYP

6. In order to not penalize schools that begin far from the state-mandated pass rate, NCLB created the "safe harbor" provision, which grants AYP to schools that fail to make AYP as described above but that reduce the number of students failing to show proficiency on the WASL by 10 percent. The safe harbor provision maintains the incentive for administrators to target the students on the margin of passing in order to show 10 percent gains. In the data this article uses, seven schools offering fourth grade, seven schools offering seventh grade, and twelve schools offering tenth grade achieved AYP through the safe harbor provision. This represents .57 percent, 1.08 percent, and 2.05 percent, respectively, of the state's elementary, middle, and high schools and .56 percent, .21 percent, and 2.05 percent of the respective state student bodies.

in consecutive years is later used to test the presence of strategic instruction. However, as Figlio and Lucas (2004) point out, schools performing poorly on state assessments affect not only themselves but also their communities through diminished property values. Thus schools face considerable incentives to improve measured performance on high-stakes tests.

In addition to the WASL, Washington students take the Iowa Test of Basic Skills (ITBS). The Iowa tests are standardized exams identifying a student's developmental level. The ITBS is given in Washington near the end of the student's third-, sixth-, and ninth-grade years, the grades prior to the WASL.[7] Using the ITBS has a number of advantages. First, since the ITBS is not employed as a tool to determine AYP, it is likely not the direct focus of strategic instruction and instead may be a tool used by administrators who decide how to allocate resources across students. Second, since the ITBS is given the year before the WASL, it can be used as a proxy for student ability. As such, this article compares the academic progress of students from the time of taking the ITBS to their completion of the WASL. Like the WASL data, the student-level ITBS mathematics scores have been normed to mean zero and unit variance within each year. Another advantage conveyed with the ITBS data is the large number of demographic, socioeconomic status (SES), and academic variables measured. These variables are used as explanatory variables in later regressions.

Optimally, a researcher would compare schools under NCLB with those that were not affected by NCLB to test whether strategic instruction took place. Yet, because all public schools in Washington are subject either to NCLB, state-level sanctions tied to NCLB, or both, there is no direct control group with which to compare strategic instruction practices. However, schools having failed to make AYP in prior years are more likely to change instruction strategies in future years in order to avoid the increasing sanctions for failing AYP. Further, both the ITBS and WASL have been given in Washington since the mid-1990s, which allows a differences-in-differences estimation strategy on the impact of NCLB. If NCLB creates strategic instructional behavior, then different WASL outcomes should be found only among schools under the threat of sanctions and should be present only after NCLB was enacted.

Data

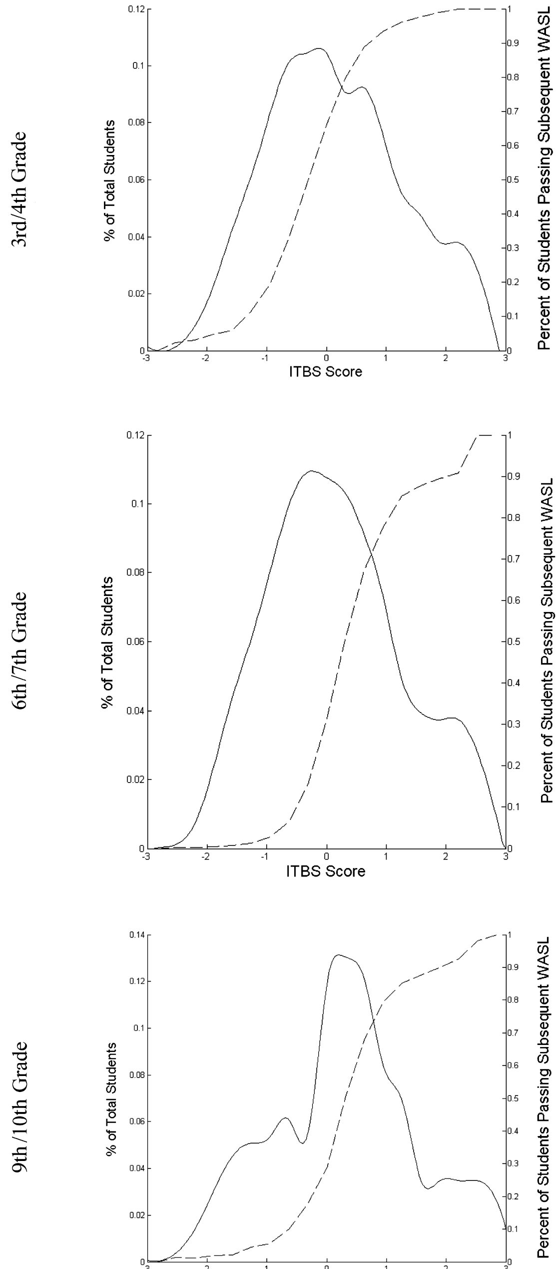
The data employed in this article consist of four cohorts of paired observations of mathematics ITBS/WASL scores for third/fourth-, sixth/seventh-, and ninth/tenth-grade students. The first observed cohorts of third, sixth, and

7. Actually, the ninth-grade test is not an ITBS but an Iowa Test of Educational Development (ITED) and covers the same areas as the ITBS in a more comprehensive fashion.

ninth graders took the ITBS in the spring of 2001 and the WASL in the spring of 2002. The final observed cohort took the ITBS in 2004 and the WASL in 2005. The state of Washington did not define AYP until late in the spring of 2002 and only notified buildings of their AYP status after the subsequent school year commenced. Hence the first two cohorts began the school year in which they took the WASL before building administrators knew their AYP status. Administrators had little ability to pursue strategic instruction for these cohorts. Students in the final two cohorts began their WASL the year after schools knew their AYP status, so administrators were in a better position to pursue strategic instruction. This heterogeneity between the two sets of cohorts presents one method of identifying the impact of NCLB.

One necessary condition for strategic instruction to occur is for building administrators to feel confident that high-ability students will demonstrate proficiency on the WASL, even if resources are diverted from those students. Likewise, administrators would be unlikely to reduce resources to low-ability students if those students had a reasonable probability of passing the WASL. The simple correlation between individual students' ITBS scores and their future binary WASL pass/no pass outcomes ranges from .55 (third/fourth grades) to .61 (sixth/seventh grades). A more revealing description of this relationship is shown in figure 1, which presents the entire sample's ITBS distribution and superimposes a plot of the fraction of students at that ITBS level who later pass the WASL. For each grade pair, virtually all students at the top ITBS scores pass the WASL. Even students scoring one standard deviation above the ITBS mean are very likely to pass the WASL. For instance, 80 percent of high school students in this situation will pass the WASL. At the same time, students scoring one standard deviation below the ITBS mean are very unlikely to pass the WASL (only 5 percent of high school students in this situation pass). Thus the opportunity for administrators to treat students on the tails as near certainties opens the possibility of diverting resources from those students to ones with a more reasonable probability of demonstrating proficiency.

As a first attempt to investigate strategic instruction, the final two cohorts are examined—the cohorts who took the WASL in buildings where the administrator knew the AYP status. After excluding special education students and those with missing observations, the pooled number of students observed in the last two cohorts at the third/fourth-grade level is 115,357; 127,196 at the sixth/seventh-grade level; and 116,376 at the ninth/tenth-grade level. This represents between 67.8 percent (sixth/seventh grades) and 68.7 percent (ninth/tenth grades) of all Washington public school students, and between 78.0 percent (ninth/tenth grades) and 89.1 percent (sixth/seventh grades) of



Dashed line represents percent of students passing WASL (right axis).

Figure 1. ITBS Distribution and WASL Proficiency

non–special education students.[8] For each grade year, students are divided into two categories: students at schools that made AYP in the previous year and students at schools that failed to make AYP in the previous year. If a building administrator fails to make AYP and reacts strategically the following year by focusing on the marginal student, the gains made by students in the tails of the ITBS distribution will be relatively smaller than those made at AYP schools.

One method of examining the strategic instruction hypothesis is to group students into quintiles based on their ITBS score and then to examine their subsequent WASL performance. Panel A in table 1 shows a comparison of student performance from the last two observed cohorts at schools that made AYP versus those at schools that did not. Consider the experience of fourth graders first. The average WASL score of fourth graders who previously scored in the lowest quintile of the ITBS and who attend schools making AYP is 1.243 standard deviations below the WASL mean. Similar students in buildings not making AYP expect to score 1.511 standard deviations below the mean, .268 standard deviations worse than their peers at AYP schools. However, this difference shrinks for students at the center of the distribution and reaches a minimum of .180 standard deviations at the fourth quintile. For students scoring in the highest quintile of the ITBS, the difference grows to over one-third of a standard deviation, indicating that high-ability students in AYP schools do significantly better than their peers at non-AYP schools. The results for seventh-grade students are similar; students on the ends of the ITBS distribution at AYP schools score more than their peers in the middle of the distribution and score significantly better than students at non-AYP schools. Interestingly, this pattern does not hold for high school students.

A number of authors demonstrate that characteristics of students at schools failing to make AYP are substantially different than those of AYP schools.[9] This accounts for the persistently higher WASL scores at AYP schools relative to non-AYP schools, observed in table 1. If student characteristics are correlated with academic growth at the tails of the ITBS distribution, then the results in table 1 may not demonstrate a strategic response of administrators to AYP

8. Dropping special education students results from the structure of NCLB. The NCLB Act requires subgroups of students, including special education students, at each building to annually make AYP. Other than special education, the composition of these groups is not determined by student ability. Since special education students are required to make AYP as a specific group and as part of a building's entire student body, it would be conceivable that administrators redirect resources within special education programs to the marginal special education student and thereby introduce special education strategic instruction in addition to the strategic instruction focused on in this article. Omitting special education students has the benefit of focusing on a source of strategic instruction within the general population of students.
9. For instance, see Krieg and Storer (2006) or Hoerandner and Lemke (2006).

Table 1. Means, Variances, and Paired t-Test Results of Student WASL Scores

		Panel A: Cohorts Where AYP Is Known				Panel B: Cohorts Where AYP Is Unknown			
	Student ITBS Quintile	Schools Making AYP in Previous Year		Schools Failing to Make AYP in Previous Year	Difference	Schools Making AYP in Future Years		Schools Failing to Make AYP in Future Years	Difference
3rd/4th	Lowest quintile	−1.243 (.714)	>	−1.511 (.749)	.268	−1.148 (.782)	>	−1.430 (.741)	.282
	Second quintile	−.646 (.621)	>	−.877 (.637)	.231	−.564 (.665)	>	−.914 (.635)	.350
	Middle quintile	−.216 (.586)	>	−.426 (.582)	.210	−.182 (.619)	>	−.475 (.609)	.293
	Fourth quintile	.202 (.601)	>	.022 (.580)	.180	.207 (.619)	>	−.051 (.600)	.258
	Highest quintile	.851 (.740)	>	.507 (.633)	.344	.843 (.732)	>	.492 (.754)	.351
6th/7th	Lowest quintile	−1.230 (.663)	>	−1.341 (.617)	.111	−1.149 (.618)	>	−1.360 (.600)	.211
	Second quintile	−.673 (.546)	>	−.563 (.548)	.110	−.562 (.548)	>	−.727 (.508)	.165
	Middle quintile	−.085 (.508)	>	−.128 (.504)	.043	−.111 (.519)	>	−.267 (.516)	.156

	Fourth quintile	.398 (.520)	>	.325 (.498)	.073	.397 (.551)	>	.182 (.521)	.215
	Highest quintile	1.105 (.643)	>	.878 (.527)	.227	1.154 (.666)	>	.896 (.615)	.258
9th/10th	Lowest quintile	−1.276 (.711)	>	−1.370 (.644)	.094	−1.199 (.674)	>	−1.336 (.622)	.137
	Second quintile	−.809 (.630)	>	−.914 (.563)	.105	−.771 (.604)	>	−.923 (.595)	.152
	Middle quintile	−.332 (.575)	>	−.497 (.562)	.165	−.331 (.558)	>	−.493 (.553)	.162
	Fourth quintile	.240 (.553)	>	.085 (.751)	.155	.208 (.572)	>	.056 (.583)	.152
	Highest quintile	.986 (.618)	>	.855 (.657)	.131	.494 (1.061)	>	.335 (1.154)	.164

Notes: Standard deviations in parentheses. > indicates the result of a paired t-test performed at the 99% level of confidence.

determination. For instance, if students of high SES are more likely to attend an AYP school and are more likely to transform a given ITBS score into a higher WASL score, then one would expect to see the differences at the upper end of the ITBS distribution as found in Panel A of table 1.

If the differences in Panel A are driven by the characteristics of schools rather than strategic decisions of administrators, the differences across ITBS performance for students prior to learning AYP status should be similar to those presented in Panel A. Panel B of table 1 sorts buildings by their future attainment of AYP and compares the performance of the first two cohorts in the data set, that is, for schools that have yet to learn their AYP status. Unlike performance in the final two cohorts at the fourth-grade level, the difference between WASL performance at schools that will make future AYP and those that will not does not change systematically across ITBS quintiles. The same is true for tenth graders, where there is no difference for the last two cohorts. The seventh-grade differences follow a pattern similar to the differences found in Panel A. Taken as a whole, table 1 suggests the presence of strategic instruction at the primary and perhaps middle school grades, but not in high schools. The next section introduces a regression model examining the impact of the ITBS on the WASL while controlling for student and school characteristics that confound the determination of strategic instruction.

Econometric Strategy and Results

To test for strategic instruction practices, consider the regression:

$$\begin{aligned} WASL_{ibt} = \sum_{j=1}^{k} \alpha_j ITBS^{j}_{ibt} + \sum_{j=1}^{k} \delta_j AYP_{bt} \times ITBS^{j}_{ibt} + \lambda AYP_{bt} \\ + \psi NCLBA_{bt} + \boldsymbol{\beta} \mathbf{X}_{ibt} + \boldsymbol{\gamma} \mathbf{Z}_{bt} + B_b + \varepsilon_{ibt} \end{aligned} \quad (1)$$

where $WASL_{ibt}$ is student i's test score in building b during time period t. $\mathbf{X}_{ibt}$ is a matrix of student-specific control variables, $\mathbf{Z}_{bt}$ represents a matrix of time-varying building control variables, and B_b and T_t represent building and time fixed effects, respectively.[10] AYP is a building-level binary variable

10. For all three grade pairs examined, the building control variables include the percent of students on free/reduced price lunch, the percent of teachers with master's degrees, average class size, and average teacher experience. Shared student-level data for all three grades examined include length of time spent at current building, gender, and ethnicity. The third/fourth- and ninth/tenth-grade regressions also include owning and using a computer at home, frequency of reading for fun, being previously held back one grade, frequency of television watching, English spoken at home, and an indicator of moving in the middle of the WASL school year. The sixth/seventh-grade regressions share measures of school attendance, illegal drug use, school spirit, frequency of parental help on homework, and perception of violence within the school with the ninth/tenth graders. Finally, the ninth/tenth graders further include measures of a student's educational goals, parents' education, athletic participation, frequency of out-of-school homework, grade point average, alcohol use, and plans after graduation. All of the building-level variables were measured during the WASL year. All

that equals one if student i's building made AYP the previous year. Since no buildings failed to make AYP prior to the passage of NCLB, the variable AYP equals one for observations in the first two periods. NCLB is a binary variable equal to one after the No Child Left Behind Act was passed and zero otherwise. ITBS is allowed to affect WASL in a nonlinear manner to allow for differential WASL gains made at various points along the ITBS distribution.

The coefficients of interest in equation 1 are the δs, which measure the impact on WASL of changes in ITBS scores by students in buildings that made AYP in the previous year. Since equation 1 employs a polynomial in ITBS, the δs potentially capture differential performance on the WASL at different levels of ITBS scores. Consider the simplest polynomial in ITBS ($k = 2$). Under the strategic instruction hypothesis, students with both high and low ITBS scores at AYP schools will score better than if their school had failed to make AYP. If true, this would result in a positive coefficient on the $AYP \times ITBS^2$ term.

Do administrators respond strategically to past AYP determination? Using quadratic and cubic polynomials in ITBS, table 2 presents regression estimates of the αs, δs, ψ, and λ for all three grade pairs observed.[11] The results are suggestive of strategic instruction in all observed grades. For instance, take the first column of table 2 as an example. This column presents results for the fourth-grade WASL with a quadratic in ITBS. The α coefficients indicate that WASL performance is increasing at a decreasing rate in ITBS performance. However, as indicated by the jointly significant δ coefficients, the impact of ITBS on WASL scores differs for students attending schools previously making AYP. Specifically, the positive quadratic term indicates a higher marginal impact on WASL scores for students at both the upper and lower ends of the ITBS distribution at schools having previously made AYP. The cubic results in the second column of table 2 reinforce these findings; at each grade level, the coefficient on the cubic interactive variable is positive, indicating growth in WASL scores for those at the upper end of the ITBS distribution. For each grade pair, figure 2 presents predicted WASL scores based on the regression results using the cubic in ITBS presented in table 2.

In order to focus on the marginal impact of making AYP, figure 2 sets all variables in equation 1, except the polynomials in ITBS, $AYP \times ITBS$, and AYP, equal to zero. The plots in figure 2 are consistent with the strategic

of the individual-level variables except previously held back one grade, gender, race, and moving during the middle of the WASL year were asked as part of the ITBS exam and are measured in the year preceding the WASL. Unreported models were added to equation 1 dummies for each time period. These were not jointly significant.

11. Quartic interactive ITBS variables were also attempted, but for all three grade levels the coefficient on these variables was statistically insignificant. The standard errors of each coefficient have been corrected for intra-group correlation of the error terms at the building level.

Table 2. WASL Fixed Effects Regression Results

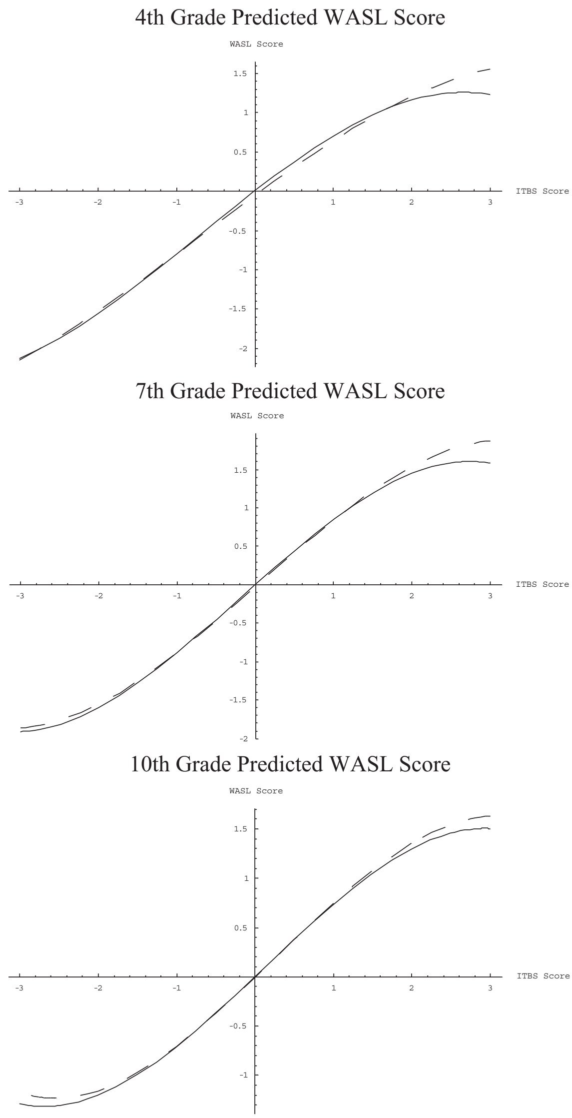
	Variable	4th WASL	4th WASL	7th WASL	7th WASL	10th WASL	10th WASL
α	ITBS	.719*** (.009)	.776*** (.015)	.803*** (.005)	.903*** (.008)	.593*** (.005)	.753*** (.007)
	ITBS^2	−.029*** (.006)	−.050*** (.007)	−.004** (.001)	−.018*** (.004)	.016*** (.003)	.012*** (.003)
	ITBS^3		−.024*** (.004)		−.035*** (.001)		−.032*** (.001)
δ	AYP×ITBS	−.017* (.009)	−.041*** (.015)	.003 (.006)	−.012 (.008)	−.00004 (.005)	.002 (.008)
	AYP×ITBS^2	.005** (.002)	.023*** (.007)	.005* (.003)	.021*** (.004)	.009** (.003)	.014*** (.003)
	AYP×ITBS^3		.011*** (.004)		.005** (.002)		.0001 (.002)
λ	AYP	−.037** (.016)	−.054*** (.017)	−.017 (.014)	−.023* (.013)	.002 (.011)	−.005 (.011)
ψ	NCLBA	−.037* (.016)	−.012** (.006)	−.0001 (.007)	−.0001 (.007)	.019** (.009)	.018** (.009)
	N	231,447	231,447	246,594	246,594	226,352	226,352
	R^2	.569	.574	.693	.699	.660	.638
	F test of $\delta = 0$	6.41**	13.50***	6.00**	9.52***	3.15**	5.85***

Notes: Standard errors corrected for clustering within buildings are in parentheses. All regressions contain building fixed effects and the independent variables listed in note 10.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

instruction hypothesis in a number of ways. First, students at the upper end of the ITBS distribution score higher on the WASL if their school made AYP. For instance, a student at an AYP school scoring 2.5 standard deviations above the mean on the ITBS in the third grade is expected to score 1.41 standard deviations above the WASL mean.[12] Had that school not made AYP, the same student would have been expected to score 1.25 standard deviations above the WASL mean, a decrease of .16 standard deviations. At the seventh grade, this predicted difference is .17 of a standard deviation, while at the tenth grade, the difference is .09 standard deviations. To put this in perspective, the (unreported) coefficient on African American (relative to white) is −.10, or about two-thirds of the difference between high-performing fourth-grade students at AYP and non-AYP schools.

The strategic instruction hypothesis suggests that students at AYP schools believed to be at the margin of demonstrating WASL proficiency should gain more than if their school had not made AYP. Table 2 presents evidence

12. 7.4 percent of the third-grade sample scored at least 2.5 standard deviations above the ITBS mean.



Dashed lines indicate results for students at AYP schools; solid lines indicate results for students at schools recently failing to make AYP.

Figure 2. Predicted WASL Scores Based on Equation 1

suggesting that this does occur. For all grade levels, the negative coefficients on AYP demonstrate that the average ITBS student expects to score better on the WASL if his or her school failed to make AYP. This difference is substantial for fourth- and seventh-grade students: .023 standard deviations for seventh-grade students and .054 standard deviations for fourth-grade students. While it need not be true that the marginal student also happens to be the student scoring at the mean of the ITBS test, these results do suggest that AYP and non-AYP schools treat their average students differently.

Under NCLB, failing to make AYP for consecutive years brings about increasing sanctions on the school and district. Thus administrators failing to make AYP for consecutive years have a greater incentive to perform strategic instruction. To test for this, consider the regression:

$$\begin{aligned} \mathrm{WASL}_{ibt} = & \sum_{j=1}^{k} \alpha_j \mathrm{ITBS}_{ibt}^{j} + \sum_{j=1}^{k} \delta_j \mathrm{AYP}_{bt} \times \mathrm{ITBS}_{ibt}^{j} + \sum_{j=1}^{k} \omega_j \mathrm{TWICE}_{bt} \\ & \times \mathrm{ITBS}_{ibt}^{j} + \lambda \mathrm{AYP}_{bt} + \upsilon \mathrm{TWICE}_{bt} + \psi \mathrm{NCLBA}_{bt} \\ & + \boldsymbol{\beta} \mathbf{X}_{ibt} + \boldsymbol{\gamma} \mathbf{Z}_{bt} + \mathrm{B}_b + \varepsilon_{ibt} \end{aligned} \qquad (2)$$

where TWICE is a binary variable equal to one if the building made AYP for the previous two consecutive years. Under this formulation, a building with TWICE equal to one also has AYP equal to one because, by definition, it made AYP in the previous year. The omitted category is buildings that failed to make AYP either in the previous year or in the previous two years.[13]

Estimated coefficients from equation 2 are presented in table 3, and predicted values from these regressions are plotted in figure 3. As this figure demonstrates, including TWICE makes little difference on the fourth- and seventh-grade results. Higher ability students in these grades perform better at AYP schools, while those at the middle of the ITBS distribution perform better at non-AYP schools. The greatest change occurred at the high school level. Students at the upper end of the ITBS distribution at high schools making AYP for two consecutive years perform far better than students at other high schools. A student attending a school making AYP for two consecutive years and who scores 2.5 standard deviations above the mean on the ITBS distribution is expected to score 1.63 standard deviations above the WASL mean.[14] A similar student at a school failing to make AYP in the previous year is expected to score 1.29 standard deviations above the mean, nearly one-quarter of a standard deviation lower than comparable peers at successful schools.

Before investigating alternative explanations for these findings, it is important to note that in both tables 2 and 3 there is mixed evidence that NCLB altered

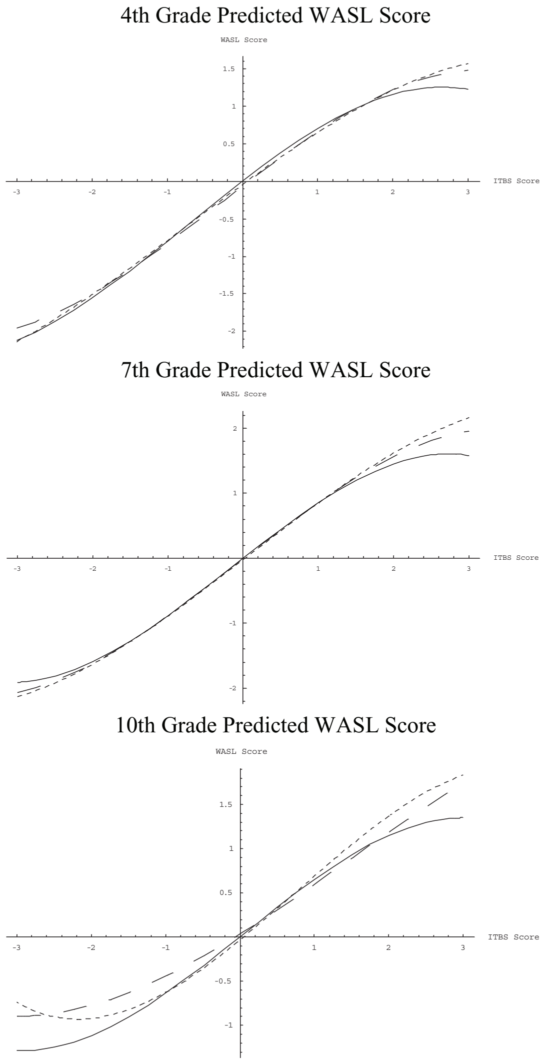
13. Four possibilities occur: buildings pass for two consecutive years, fail for two consecutive years, pass in the previous year while failing two years ago, or fail in the previous year while passing two years ago. The formulation of equation 2 merges the failing for two consecutive years with failing in the previous year and passing two years ago. An unreported set of regressions verified that this merger results in no substantive differences in the reported results when compared to a regression controlling for all four possibilities.
14. 8.6 percent of high school students scored 2.5 standard deviations above the WASL mean.

Table 3. WASL Fixed Effects Regression Results

	Variable	4th WASL	7th WASL	10th WASL
α	ITBS	.776*** (.015)	.903*** (.008)	.669*** (.007)
	$ITBS^2$	−.050*** (.007)	−.018*** (.004)	.003 (.003)
	$ITBS^3$	−.024*** (.004)	−.035*** (.001)	−.025*** (.001)
δ	AYP×ITBS	−.008 (.020)	−.007 (.012)	−.140*** (.012)
	AYP×$ITBS^2$	.030*** (.010)	.013** (.006)	.040*** (.006)
	AYP×$ITBS^3$	.003 (.006)	.010*** (.003)	.015*** (.003)
ω	TWICE×ITBS	−.034* (.017)	−.005 (.011)	.158*** (.010)
	TWICE×$ITBS^2$	−.007 (.008)	.009** (.004)	.061*** (.005)
	TWICE×$ITBS^3$	.009** (.004)	.005** (.002)	−.018*** (.003)
λ	AYP	−.067** (.025)	−.018* (.010)	.029* (.015)
υ	TWICE	.022 (.026)	−.018 (.017)	−.094*** (.017)
ψ	NCLBA	−.027*** (.007)	−.001 (.007)	.009 (.010)
	N	231,447	246,594	226,352
	R^2	.575	.700	.665
	F test of $\delta = 0$	10.84**	6.08***	77.64***
	F test of $\omega = 0$	6.34**	6.36***	153.74***

Notes: Standard errors corrected for clustering within buildings are in parentheses. All regressions contain building fixed effects and the independent variables listed in note 10.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

overall WASL performance. In both tables, a positive impact of NCLB is found on tenth-grade students (although this coefficient is insignificant in table 3). A potential explanation for this gain is that during the post-NCLB years, the state of Washington required high schools to post WASL scores on student transcripts. Unlike tenth graders, NCLB affected fourth- and seventh-grade scores negatively. While a number of reasons for this might be suggested, it is possible that for all grades involved, NCLB reduced the efficiency of instruction to all students, and only secondary students, with the external pressure of having scores placed on transcripts, overcame this.



The frequently dashed line represents students at schools making AYP for consecutive years; the infrequently dashed line represents students at schools making AYP just in the previous year; the solid line represents students at schools either failing to recently make AYP or failing for two consecutive years.

Figure 3. Predicted WASL Scores Based on Equation 2

Alternative Explanations and Robustness Checks

One possible explanation for these findings is that omitted interactions between ITBS scores and AYP bias the δs in such a way as to lead to the conclusion of strategic instruction. For instance, if schools making AYP inherently transform the ITBS scores into WASL scores differently than non-AYP schools and this difference is not accounted for by the building fixed effects, then the

Table 4. WASL Fixed Effects Regression Results

	Variable	4th WASL	7th WASL	10th WASL
α	ITBS	.746*** (.012)	.892*** (.008)	.742 (.008)
	$ITBS^2$	−.035*** (.005)	.017*** (.003)	.041 (.002)
	$ITBS^3$	−.019*** (.003)	−.031*** (.001)	−.034 (.001)
δ	Future AYP×ITBS	−.038*** (.013)	.001 (.009)	−.008 (.009)
	Future AYP×$ITBS^2$	.006 (.006)	−.004 (.004)	−.004 (.003)
	Future AYP×$ITBS^3$	.002 (.003)	−.0007 (.002)	.0001 (.002)
	N	116,090	121,398	109,981
	R^2	.549	.693	.691
	F test of $\delta = 0$	3.58**	.48	1.11

Notes: Standard errors corrected for clustering within buildings are in parentheses. All regressions contain building fixed effects and the independent variables listed in note 10.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

δs may be misinterpreted as strategic instruction. To check for this possibility, consider the regression

$$WASL_{ibt} = \sum_{j=1}^{k} \alpha_j ITBS^j_{ibt} + \sum_{j=1}^{k} \delta_j \text{Future AYP}_{bt} \times ITBS^j_{ibt} + \boldsymbol{\beta} \mathbf{X}_{ibt} + \boldsymbol{\gamma} \mathbf{Z}_{bt} + B_b + \varepsilon_{ibt} \quad (3)$$

where Future AYP is equal to one for each student attending a building that will make AYP in each of the two years after enactment of NCLB. This regression is analogous to panel B of table 1 and is applied only to pre-NCLB observations. Since no motive existed for strategic instruction prior to NCLB, the δs in this regression should be zero unless unobserved inherent differences between future AYP and future non-AYP schools exist.

Estimates from equation 3 are provided in table 4. For seventh and tenth graders, the estimated δs from this equation are individually and jointly insignificantly different than zero, indicating that schools that will make AYP did not transform student ITBS scores into WASL scores differently than non-AYP schools prior to NCLB. For fourth graders, only the interaction of Future AYP×ITBS had a significant negative sign, indicating that elementary schools that would make AYP actually did less well transforming high ITBS scores into WASL scores prior to NCLB. For all three grades, there is no evidence to suggest that the earlier strategic instruction findings are driven by unobserved characteristics that were present prior to NCLB.

A second explanation for these findings is that the composition of students taking the exam differs between AYP and non-AYP schools and this difference is not accounted for by the independent variables used in the regressions. This concern has been addressed by a number of studies, especially with regard to school placement of students into special education programs.[15] However, the special education argument is unlikely to affect the finding of strategic instruction. The data presented above omit students in special education programs. If schools in danger of sanctions under NCLB are more likely to assign students to special education, then the conditional WASL performance for remaining students should be higher at those schools than that at AYP schools. However, in figures 1 and 2, there is only one case in which non-AYP students at the bottom of the ITBS distribution outperform those at AYP schools (seventh grade, figure 3), suggesting that if schools strategically place students into special education programs, this action has little impact on conditional WASL scores.[16]

A more insidious form of selection bias occurs because some students are excluded from the regression as a result of missing observations of test scores. Table 5 presents counts of missing observations by grade and year. Over the time period examined, the percentage of valid general education students observed has increased for each grade. Much of this increase can be attributed to a greater fraction of students taking the WASL. Because NCLB requires schools to test at least 95 percent of their enrolled students, this pattern is not surprising.[17] However, it is possible that students missing the WASL do so for reasons correlated with their ability and school AYP determination. For instance, it may be the case that schools purposely encourage high-ability students to take the WASL while discouraging low-ability students to do so. The incentive to do this would be stronger in non-AYP schools in order for them to regain AYP status. As a matter of fact, Figlio (2006) finds that during test weeks in Florida, the duration of disciplinary suspensions for low-performing students in grades that face high-stakes tests increases. If nonrandom selection

15. See Figlio and Getzler (2002), Deere and Strayer (2001), Cullen and Reback (2006), and Jacob (2005).
16. This concern is also minimized by the NCLB requirement that 95 percent of a building's students must be tested, leaving administrators with little room for manipulating the composition of students taking high-stakes exams.
17. A second pattern arising in table 5 is the large decrease in valid observations that occurs in the tenth grade. This increase is almost completely driven by tenth graders missing the WASL, which, in turn, is most likely a function of dropping out of school. The statewide dropout rate for ninth graders (some of whom took the ITBS test in the ninth grade) was 4.1 percent in 2004–5, and the dropout rate for tenth graders was 4.3 percent.

Table 5. Numbers of Included and Excluded Students

Academic Year	General Education	Missing ITBS Score	Missing WASL Score	Missing Both ITBS & WASL	Valid Observations
a. Grade 4					
2001–2	67,346	4,837 (7.1%)	3,742 (5.5%)	814 (1.2%)	57,953 (86.1%)
2002–3	67,878	5,231 (7.7%)	3,876 (5.7%)	634 (.9%)	58,137 (85.6%)
2003–4	65,583	4,674 (7.1%)	3,440 (5.2%)	656 (1.0%)	56,813 (86.6%)
2004–5	65,669	3,709 (5.6%)	2,905 (4.4%)	511 (.7%)	58,544 (89.1%)
Total	266,476	18,451 (6.9%)	13,963 (5.2%)	2,615 (1.0%)	231,447 (86.8%)
b. Grade 7					
2001–2	70,128	5,501 (7.8%)	4,663 (6.6%)	1,028 (1.4%)	58,936 (84.0%)
2002–3	72,171	4,973 (6.9%)	4,008 (5.5%)	728 (1.0%)	62,462 (86.5%)
2003–4	72,307	5,479 (7.5%)	4,447 (6.1%)	907 (1.2%)	61,474 (85.0%)
2004–5	70,395	3,301 (4.7%)	2,827 (4.0%)	545 (.8%)	63,722 (90.5%)
Total	285,001	19,254 (6.7%)	15,945 (5.6%)	3,208 (1.1%)	246,594 (86.5%)
c. Grade 10					
2001–2	73,725	6,624 (9.0%)	8,756 (11.9%)	2,734 (3.7%)	55,611 (75.4%)
2002–3	72,820	6,785 (9.3%)	8,785 (12.1%)	2,885 (4.0%)	54,365 (74.6%)
2003–4	74,006	6,745 (9.1%)	8,525 (11.5%)	2,757 (3.7%)	55,979 (75.6%)
2004–5	75,113	5,303 (7.0%)	6,856 (9.1%)	2,557 (3.4%)	60,397 (80.4%)
Total	295,664	25,457 (8.6%)	32,922 (11.1%)	10,933 (3.7%)	226,352 (76.6%)

Note: Numbers in parentheses represent the percent of total general education students.

of students occurs, then the results of tables 3 and 4 may be biased in favor of finding strategic instruction.

To test for the possibility of sample selection bias, a two-stage Heckit procedure is employed.[18] In the first stage, a probit uses the regressors from equation 1 and the contemporaneous percentage change of a county's population to estimate if a student missed the WASL. Because a primary reason for missing WASL observations is that students move from their local school district, including the percentage change in the local population may help explain attrition from the sample. The second stage includes the inverse Mills ratio from this Probit in equation 1. For each grade pair, results from the first- and second-stage regressions are presented in table 6.

18. See Wooldridge 2002, chapter 17, for details.

Table 6. First- and Second-Stage Heckit Results

	Variable	Grade 4 1st-Stage Probit	4th WASL	Grade 7 1st-Stage Probit	7th WASL	Grade 10 1st-Stage Probit	10th WASL
α	ITBS	−.020 (.062)	.801*** (.025)	−.759*** (.081)	.902*** (.018)	−.310*** (.040)	.614*** (.007)
	ITBS^2	.019 (.037)	−.075*** (.020)	.053 (.050)	−.018*** (.004)	.017 (.024)	.007** (.003)
	ITBS^3	−.025 (.020)	.008 (.025)	.027 (.025)	−.035*** (.002)	.010 (.012)	−.023* (.013)
δ	AYP×ITBS	−.138** (.064)	.135** (.050)	.166* (.086)	−.011 (.008)	−.055 (.045)	.011 (.009)
	AYP×ITBS^2	.036 (.037)	−.022 (.036)	.040 (.052)	.021*** (.004)	.004 (.027)	.015*** (.004)
	AYP×ITBS^3	.002 (.020)	.009** (.004)	−.034 (.026)	.005** (.002)	.002 (.014)	−.001 (.002)
λ	AYP	−.175*** (.056)	.171 (.175)	−.303*** (.081)	−.023 (.016)	−.121*** (.045)	−.023* (.013)
ψ	NCLBA	−.116*** (.016)	.134 (.114)	−.360*** (.042)	−.0003 (.010)	−.287*** (.034)	−.020 (.015)
	% county pop. growth	.016* (.010)	—	.012* (.007)	—	.183*** (.027)	—
	Inverse Mills ratio		1.28 (.995)		−.0005 (.053)		−.163*** (.051)
	N	245,410	231,447	262,539	246,594	259,274	226,352
	R^2	—	.575	—	.699	—	.663
	F test of $\delta = 0$	—	10.08**	—	9.40***	—	8.75***

Notes: Each probit regression estimates a random effects model with dependent variable equal to one if the observation missed the WASL score and zero otherwise. Standard errors corrected for clustering within buildings are in parentheses.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

Analysis of the first-stage probits in table 6 reveals two consistent findings across grades. First, schools having made AYP in the previous year are less likely to have students miss the WASL in the current year. Perhaps parents are supportive of successful schools and are more willing to enroll students and ensure their daily attendance at these schools. Of course it may also indicate that successful schools provide incentives for students to take the WASL. Second, the percent change in county population growth is a positive predictor of missing the WASL and, although significant at only the 10 percent level for fourth and seventh grades, it is highly significant for high school students. Interestingly, there is mixed evidence relating missing the WASL to a student's ITBS score. For fourth graders, the ITBS score was not a predictor of missing the WASL, while the interaction of AYP and ITBS was. For these students, the negative coefficient on the interaction of AYP and ITBS suggests that better students at AYP schools are more likely to take the WASL, exactly what one would expect if schools purposely encouraged more able students to take the WASL. However, a positive coefficient appears on this same variable in the seventh-grade probit, suggesting that more able seventh graders are less likely to take the WASL. Interestingly, for tenth graders, no differential WASL attrition based on the interaction of AYP and ITBS is demonstrated.

A comparison of the second-stage Heckit results presented in table 6 with those of the original model shown in table 2 reveals a number of important findings. First, only for high school students is the inverse Mills ratio statistically significant, indicating that WASL attrition has little impact on the fourth- and seventh-grade results. This is confirmed by the relatively small changes to the ITBS and AYP$\times$ITBS coefficients for those grades relative to the results in table 2. Second, for all grades, the coefficients on AYP$\times$ITBS are jointly significant and continue to follow the pattern described in table 2—that is, the coefficient on AYP$\times$ITBS is negative and the coefficients on the higher order interactive terms are positive. Whereas the tenth-grade results do include a statistically significant coefficient on the inverse Mills ratio, the inclusion of the Mills ratio actually results in an increase in the AYP$\times$ITBS^2 coefficient, suggesting that not accounting for selection of students who take the WASL actually causes an understatement of the strategic instruction findings. Finally, for all grades, the attrition-corrected coefficients on the binary variables AYP and NCLB were measured with less precision and/or the coefficients moved toward zero, so only the variable AYP was significantly less than zero (at the 10 percent level) for fourth and tenth grades.

Despite little evidence that WASL attrition has driven the strategic instruction findings, it is possible that nonrandom attrition from the ITBS test has led to these findings. To address ITBS attrition, a second two-stage procedure is followed. The first stage employs the subsample with complete observations of ITBS scores and performs the regression

$$\mathrm{ITBS}_{ibt} = \beta X_{ibt} + \gamma Z_{bt} + B_b + \varepsilon_{ibt}. \tag{4}$$

Using the estimated coefficients from equation 4, predicted ITBS scores are generated only for those students with missing ITBS scores.[19] The students with generated ITBS scores are then integrated into the sample and equation 1

19. Students having missing values of the ITBS were also likely to have missing observations of X in equation 3. In these cases, after estimating equation 3 using the observations without missing data, the sample average values of X were inserted for those individuals with missing observations of X in order to generate the predicted value of ITBS. This method had no substantive impacts on the strategic instruction findings.

Table 7. WASL Fixed Effects Regression Results Replacing Missing ITBS Scores with Generated ITBS Scores

		Panel A			Panel B		
	Variable	**4th WASL**	**7th WASL**	**10th WASL**	**4th WASL**	**7th WASL**	**10th WASL**
α	ITBS	.754*** (.016)	.860*** (.010)	.646*** (.008)	.808*** (.029)	.810*** (.020)	.614*** (.018)
	$ITBS^2$	−.034*** (.007)	.0008 (.004)	.011*** (.003)	−.076*** (.020)	.0004 (.011)	.006** (.003)
	$ITBS^3$	−.021*** (.004)	−.028*** (.002)	−.025*** (.001)	.013 (.029)	−.019** (.010)	−.023*** (.001)
δ	AYP×ITBS	−.036** (.017)	.008 (.010)	−.001 (.008)	.041 (.085)	.002 (.020)	−.011 (.009)
	AYP×$ITBS^2$	.016** (.007)	.017*** (.004)	.013*** (.004)	.025 (.039)	.013*** (.005)	.016*** (.003)
	AYP×$ITBS^3$	.009* (.004)	.0005 (.002)	.003 (.002)	.006 (.010)	.0004 (.004)	.002 (.002)
λ	AYP	−.086*** (.018)	−.073*** (.016)	.009 (.010)	.250 (.236)	−.043 (.032)	−.023 (.013)
ψ	NCLBA	−.012* (.006)	.0004 (.007)	.024*** (.009)	.172 (.143)	.0003 (.009)	−.020 (.015)
	N	249,898	265,848	251,809	249,898	265,848	251,809
	R^2	.505	.607	.601	.556	.699	.641
	F test of $\delta = 0$	8.03**	5.04***	4.21***	9.94**	9.58***	8.75***

Notes: Standard errors corrected for clustering within buildings are in parentheses. All regressions contain building fixed effects and the independent variables listed in note 10.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

is reestimated. The estimated regressions from this sample are presented in panel A of table 7.

Because generated ITBS values for those students with original missing values are measured with error, the ITBS and ITBS×AYP coefficients presented in panel A of table 7 will be biased toward zero.[20] As a matter of fact, when comparing these coefficients in tables 2 and 6, all are attenuated toward zero. However, for each grade, the coefficients on the ITBS/AYP interactive variables are jointly significant and maintain the same pattern as the earlier results. In short, it appears that ITBS attrition is not the cause of finding strategic instruction.

As a final check for sample selection bias, the preceding analyses were merged. Using equation 4, ITBS scores were created for those individuals with missing ITBS observations and integrated into the data. Using these data, a

20. See Greene 2003, pp. 83–90.

second Heckit procedure accounting for the missing WASL observations was estimated. Second-stage results from this Heckit are produced in Panel B of table 7. For all grade levels, δ is jointly significant and positive for higher orders of ITBS scores. While many individual δs are not statistically significant, this is likely the result of a combination of coefficients attenuated toward zero and higher standard errors generated by the inclusion of the inverse Mills ratio in the second stage of the Heckit model.

A second robustness check involves trimming the sample to control for school outliers. The possibility exists that, based on the composition of their student bodies, some schools are either so certain to make AYP or so certain to not make AYP that administrators face no incentive to perform strategic instruction. If this is the case, then the results presented actually understate the impact of strategic instruction within those schools that perform it. On the other hand, there is high variance across schools in measures like free and reduced price lunch participation, academic achievement of teachers, student demographics, and resources per pupil. If the decision to participate in strategic instruction is correlated with these measures, it is possible that a few schools acting as outliers lead to the strategic instruction conclusion.

To trim the sample, consider the building-level logit regression:

$$\Pr(Y_b = 1) = \beta \mathbf{Z}_b + D_b + \varepsilon_i \tag{5}$$

where Y is equal to one if a building fails to make AYP in either 2004 or 2005, Z represents a matrix of building-level control variables measured in 2002, and D represents a district-level fixed effect.[21] This regression can be viewed as a forecast of which schools will fail to make AYP, based on their characteristics measured at the time NCLB was enacted. From this logit regression, predicted probabilities of failing to make AYP for each school are generated and sorted. Students are then trimmed from the sample if they attended a building in either the top or bottom deciles of predicted probability of failure. Using this trimmed sample, equation 1 is reestimated, with the results presented in table 8.

For all grades, the results of the trimmed sample closely match those of table 2. The coefficients on the interactive AYP and ITBS variables are jointly significant and follow the same pattern as those in table 2. The strategic instruction conclusion holds for a more homogenous set of schools, as well as for the entire population of elementary and middle schools.

21. The building-level control variables include the percentage of students in the free or reduced price lunch program, the percentage of teachers achieving a master's degree, six variables indicating the percentage of students of a particular racial background, average class size, average teacher experience, total building size, and seven binary variables indicating the rural or urban nature of the building.

Table 8. Trimmed Regression Sample

	Variable	Grade 4	Grade 7	Grade 10
α	ITBS	.781*** (.017)	.885*** (.009)	.644*** (.009)
	ITBS^2	−.054*** (.008)	−.014*** (.004)	−.001 (.003)
	ITBS^3	−.025*** (.004)	−.030*** (.002)	−.021*** (.002)
δ	AYP×ITBS	−.043** (.017)	.008 (.009)	.012 (.011)
	AYP×ITBS^2	.027*** (.008)	.027*** (.005)	.019*** (.004)
	AYP×ITBS^3	.011** (.005)	.003* (.002)	−.002 (.002)
λ	AYP	−.058*** (.020)	−.023 (.018)	−.010 (.015)
ψ	NCLBA	−.015** (.007)	.009 (.011)	.017 (.011)
	N	190,854	129,762	137,378
	R^2	.571	.694	.661
	F test of $\delta = 0$	12.73***	3.71**	6.94***

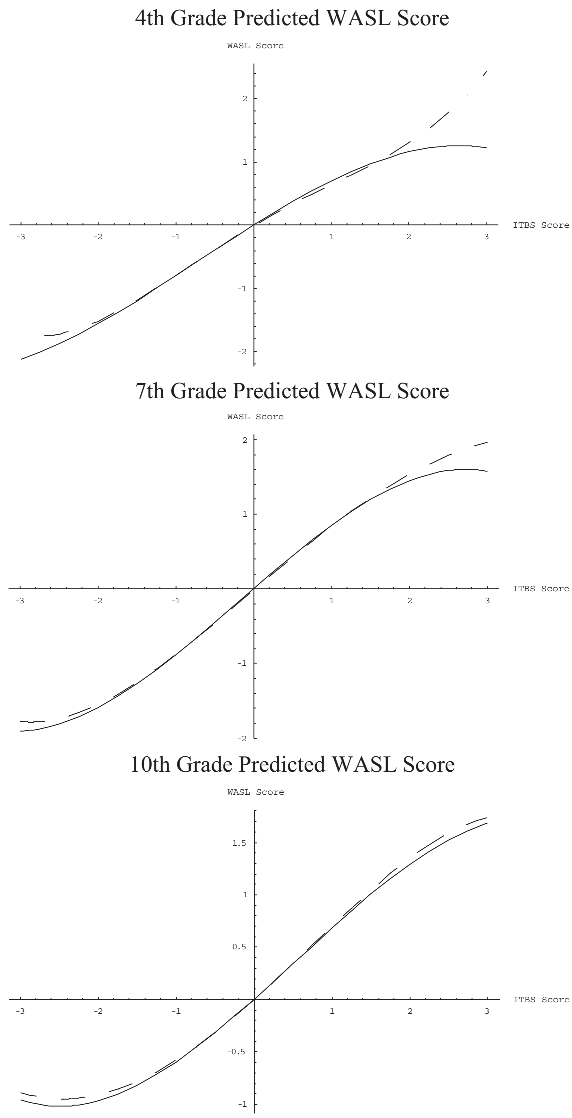
Notes: Standard errors corrected for clustering within buildings are in parentheses. All regressions contain building fixed effects and the independent variables listed in note 10.
* statistically significant at 10%; ** statistically significant at 5%; *** statistically significant at 1%.

A final concern is that the polynomial regressions used impose a structure on the results that incorrectly generates the findings of strategic instruction. To check for this possibility, consider a variant to equation 1:

$$WASL_{ibt} = f(ITBS_{ibt}) + \beta X_{ibt} + \gamma Z_{bt} + B_b + T_t + \varepsilon_{ibt}. \quad (6)$$

In this case, f($ITBS_{ibt}$) is estimated using a kernel regression.[22] The benefit of employing this estimator is flexibility; a priori, no structural relationship between the ITBS and WASL is assumed. Rather, this technique allows the data to determine the shape of the regression function, rather than requiring the researcher to employ limiting conditions that may predetermine the outcome of the regression. Since the variable of interest is the impact of previously making AYP, the kernel regression is estimated separately for schools making AYP in the previous year and those not making AYP. Further, because the first

22. Härdle (1990) details the exact procedure of kernel regressions. DiNardo and Tobias (2001), in a broad overview, describe the process employed in this article.



Dashed lines indicate results for students at AYP schools; solid lines indicate results for students at schools recently failing to make AYP.

Figure 4. Predicted WASL Scores Based on Kernel Regressions

two years of observations in this sample occur prior to schools knowing their AYP status, the kernel regression is applied to only the final two years of the sample.

Graphical results from the kernel regression appear in figure 4. Comparing these to figure 2, the results generated from equation 1, a high degree of similarity in results is revealed across techniques. For all three grades, the pattern of both high- and low-ability students at AYP schools outperforming their peers in non-AYP schools holds. As a matter of fact, under the kernel regressions, this difference is even larger at the lower end of the ITBS distribution than was found in figure 2. Further, for fourth-grade results, the differences at the upper end of the ITBS distribution appear larger in the kernel regression than the parametric ones. Thus it appears that the regression technique chosen does not affect the conclusion of strategic instruction and, if anything, the parametric results may mute the impact of strategic instruction on the tails of the ITBS distribution.

3. DISCUSSION AND CONCLUSIONS

This article demonstrates a differential impact of NCLB on students based on student ability. Specifically, because NCLB provides incentives to focus on the marginal student, students on the tails of the ability distribution gain less at schools in immediate danger of being sanctioned under NCLB. The size of these losses for elementary students in schools failing to make AYP are large and growing for students at schools that repeatedly fail to make AYP. These differences did not occur prior to enactment of NCLB and occur in the presence of building-specific fixed effects, suggesting that these results are not driven by time-invariant unobservable characteristics. These results are also robust to sample selection corrections and data trimming.

An argument may be made that these findings underestimate the true impact of NCLB. Consider a school that marginally made AYP in the previous year. Because the required pass rate rises each year, should this school fail to increase the percentage of students passing, it would fail to make AYP in subsequent years. A school in this position would have incentives to perform strategic instruction but, under the techniques employed in this research, would have been considered an AYP school. Thus the estimated impacts of strategic instruction may actually understate the actual impacts of this resource shifting.

While a movement of resources from students at the tails of the ability distribution to those in the center may be unappealing from a parental standpoint, it is not obvious that this behavior is an inefficient use of resources. For instance, if, prior to NCLB, schools overemphasized learning for students on these tails, then the decision of administrators to focus on the marginal student may redirect resources to a more efficient use. On the other hand, it is possible to evaluate school performance in a way that does not provide as strong an incentive to redirect resources from students at the tails. For example, rather than evaluating schools using a required pass rate, the state could measure individual student WASL scores conditional on prior test scores. Students could be grouped according to prior ability (perhaps highly capable learners, "average" learners, etc.) and then the schools could be evaluated on the growth of students within each ability group.

The ramifications of the current incentive structure can, in the future, be significant for both students and schools. By neglecting students in the lower tail of the ability distribution, schools may raise their chance of making AYP this year but fail to make AYP in the future, as the required pass rate converges to 100 percent and their low-ability students have been previously poorly prepared to succeed. This may be especially problematic in districts where students advance to different buildings as they age. For instance, a middle school principal may decide to strategically instruct, knowing that it will allow his school to make AYP while reducing the probability that his students will demonstrate proficiency at the high school level. Further, if the most talented students are also those who have more opportunities to be educated outside of public schools, the reduction in attention these students receive may drive them from public schools and further reduce the chances of making future AYP.

A second concern regarding strategic instruction has to do with one of the core concerns with NCLB: racial disparities. The schools most likely to fail to make AYP in Washington are those that serve minorities, low-income areas, and urban centers. By directing fewer resources toward their most talented students, these schools lower the future probability of academic and job market successes of the very minorities that NCLB proposes to help.

This article sidesteps two issues that require further investigation. First, AYP determination is based not only on overall student performance but also on the performance of eight demographic subgroups. A school with too many Hispanics that cannot achieve proficiency on the high-stakes exam will not make AYP, even if the percent of all students showing proficiency is above the required pass rate. Under the strategic instruction hypothesis, this design can lead to focusing attention both within and across demographic groups. Potentially, this type of resource shifting could increase learning for students in a particular demographic group at the expense of those in the general population. While these actions will be captured by the above analysis, more work is needed to determine if strategic instruction leads to resource shifting across demographic subgroups.

The second issue has to do with the minimum size of the demographic group required to determine non-AYP. Under NCLB, schools with fewer than thirty students in any demographic group automatically receive AYP for that group. A school with fewer than thirty students in the federally proscribed demographic group will automatically make AYP in that demographic group. While this reduces the incentives to strategically instruct at small schools, it does introduce other dimensions for strategic response. For instance, school districts may change attendance boundaries so no single school has more than thirty students in a demographic group at a particular school. As a matter of fact, some school programs that target a particular group of students—for instance, English as a second language—might be moved from one school to another each year so as to reduce the probability of failing to make AYP in multiple successive years. This incentive may also encourage districts to open additional schools in order to reduce enrollment of demographic groups below the federal threshold.

Generous funding for this research was provided through the Washington Educational Research Association and Western Washington University's Bureau for Faculty Research. I am grateful to Paul Storer, Vinit Jagdish, Mark Stater, Grant Forsyth, Julie Hansen, Rick Thompson, Yvonne Durham, and participants at the 2006 Southern Economic Association meetings for comments that improved this work. Pete Bylsma and Lisa Ireland were generous and extremely helpful with confidential OSPI data.

REFERENCES

Burgess, Simon, Carol Propper, Helen Slater, and Deborah Wilson. 2005. Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools. Working Paper No. 05/128, University of Bristol.

Chakrabarti, Rajashri. 2007. Vouchers, public school response, and the role of incentives: Evidence from Florida. Staff Report No. 306, Federal Reserve Bank of New York.

Cullen, Julie Berry, and Randall Reback. 2006. Tinkering toward accolades: School gaming under a performance accountability system. NBER Working Paper No. 12286.

Deere, Donald, and Wayne Strayer. 2001. Putting schools to the test: School accountability, incentives, and behavior. Working Paper No. 113, Texas A&M University.

DiNardo, John, and Justin L. Tobias. 2001. Nonparametric density and regression estimation. *Journal of Economic Perspectives* 15: 11–28.

Figlio, David N. 2006. Testing, crime and punishment. *Journal of Public Economics* 90: 837–51.

Figlio, David N., and Lawrence S. Getzler. 2002. Accountability, ability, and disability: Gaming the system. NBER Working Paper No. 9307.

Figlio, David N., and Maurice Lucas. 2004. What's in a grade? School report cards and the housing market. *American Economic Review* 94: 591–604.

Greene, William H. 2003. *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.

Härdle, Wolfgang. 1990. *Applied nonparametric regression*. New York: Cambridge University Press.

Hoerandner, Claus M., and Robert J. Lemke. 2006. Can No Child Left Behind close the gaps in pass rates on standardized tests? *Contemporary Economic Policy* 24: 1–17.

Jacob, Brian. 2005. Accountability, incentives, and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics* 89: 761–96.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Stecher. 2000. What do test scores in Texas tell us? Rand Issue Paper No. IP-202. Available www.rand.org/pubs/issue_papers/2006/IP202.pdf. Accessed 15 November 2007.

Krieg, John M., and Paul Storer. 2006. How much do students matter? Applying the Oaxaca decomposition to explain determinants of adequate yearly progress. *Contemporary Economic Policy* 24: 563–81.

Ladd, Helen F. 2001. School-based educational accountability systems: The promise and the pitfalls. *National Tax Journal* 54: 385–400.

Reback, Randall. 2008. Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*. In press.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.