Economics of Education Review xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Which students are left behind? The racial impacts of the No Child Left Behind Act☆

John M. Krieg*

Western Washington University, MS 9074, Department of Economics, Bellingham, WA 98225, United States

ARTICLE INFO

Article history:
Received 24 July 2009
Received in revised form 7 February 2011
Accepted 9 February 2011

JEL classification:
I28

Keywords:
Strategic instruction
Education incentives

ABSTRACT

The No Child Left Behind Act imposes sanctions on schools if the fraction of any of five racial groups of students demonstrating proficiency on a high stakes exam falls below a statewide pass rate. This system places pressure on school administrators to redirect educational resources from groups of students likely to demonstrate proficiency towards those who are marginally below proficient. Using statewide observations of 3rd and 4th grade math tests, this paper demonstrates that students of successful racial groups at schools likely to be sanctioned gain less academically over their subsequent test year than comparable peers at passing schools. This effect is stronger at schools more likely to suffer from NCLB sanctions and is robust to corrections for non-random sample selection.

1. Introduction

Demands for school accountability and concerns about racial performance disparities culminated in the No Child Left Behind Act (NCLB), the 2002 reauthorization and expansion of the Elementary and Secondary School Act. NCLB holds districts and buildings accountable for student performance on state administered high-stakes tests, sanctions failing schools, and provides expanded educational opportunities for students attending these schools. Proponents of NCLB hope it will increase educational quality and reduce the racial and income academic achievement gaps. However, the implementation of NCLB inadvertently provides incentives to reduce academic achievement for some groups of students. This article describes these incentives and documents a reduction in scholastic performance among these groups.

NCLB institutes a system of performance goals that, if not met, trigger sanctions of increasing severity on schools and districts. Yet, as Campbell's Law suggests, "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."[1] Specifically, NCLB creates incentives for school administrators to focus resources on specific subgroups of students in the hopes of making Adequate Yearly Progress (AYP). As mandated by NCLB, each school must test five distinct racial groups and three categories of students: Black, Hispanic, White, American Indian, Asian/Pacific Islander, low-income, bilingual and special education. For a school to make AYP, the percentage of students in each group within that school who demonstrate proficiency on a high-stakes test must exceed a state determined pass rate. Schools with too low a percent of students demonstrating proficiency in any group do not make AYP and are subject to school-wide sanctions under NCLB. By focusing on a binary pass/no pass outcome for each group, NCLB provides incentives for administrators

☆ This paper was generously funded by the Washington Educational Research Association. Thanks to Allan Sleeman for helpful comments. All remaining errors are my own.

* Tel.: +1 360 650 7405.
E-mail address: John.Krieg@wwu.edu

[1] Campbell (1976, p. 49).

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.02.004

to direct resources away from groups projected to make AYP and target those resources towards members of groups thought to be in danger of not making AYP. For instance, an administrator may choose to abandon a curriculum that has broad appeal for one that focuses on skills that a lower performing group of students lack. Administrators may assign students of weaker groups to stronger teachers in hopes of raising their high-stakes academic performance leaving students of other groups in the care of less able teachers. Class sizes could be adjusted to benefit weaker students, administrators may choose to fund co-curricular activities that appeal to one particular group in hopes of raising their academic performance, or teachers may divert attention from low-stakes subjects to high-stakes subjects (Winters, Trivitt, & Greene, 2010) which may have differential impacts on demographic subgroups. Brown and Clift (2010) present evidence that teachers abandon advanced material in favor of more basic material in hopes of aiding failing students which promotes a "curricular ceiling"[2] on proficient students. They also document evidence that one school in their study provided each teacher with a list of 5 students in their class who, if demonstrating proficiency on a high stakes test, would allow the school to make AYP. Teachers involved claimed that this list focused attention on those 5 students at the expense of others in their class. Rouse, Hannaway, Goldhaber, and Figlio (2007) document Florida schools who failed that state's accountability standards were more likely to reorganize students within classrooms into smaller learning "units", were more likely to mandate a minimum class time spent on high stakes subjects, and were more likely to reward high teacher performance. In a survey of principals, Stecher et al. (2008) found that a clear majority of principals reported encouraging teachers to "focus their efforts on students close to meeting the standards."[3] Whatever the specific avenue, responding to the possibility of failure under NCLB in this way is referred to as "strategic instruction" in this paper. This paper focuses on racial groups as the target for strategic instruction. The racial focus is driven by the fact that the state of Washington releases all needed information on a student's race but does not provide data on two of the other three NCLB subgroups: individual student income or bilingual status. The third subgroup, special education students, because of their wide range in abilities, testing techniques, and expected outcomes, are excluded from the analysis.

To demonstrate the identification strategy used to evaluate strategic instruction, consider two similar students. The first is a member of a racial group which made AYP in the prior year but attends a school that contains another racial group that failed to make AYP. The second is a member of the same racial group as the first but attends a school that had no groups fail to make AYP. If strategic instruction exists, then the first student should gain less academically over the course of the subsequent year than the second because resources are directed away from the first student in favor of the failing racial group at his or her school. Measuring academic differences between these students suggests one method of identifying strategic instruction. The data employed in this paper presents another identification dimension. The data span the period before and after enactment of NCLB presenting the ability to measure the difference in the difference between these students that occurred before and after NCLB.

This paper presents econometric estimates that are consistent with the strategic instruction hypothesis. Using a statewide sample of 4th graders, it is found that students of successful racial groups who attend schools where another racial group fails to make AYP score lower on a subsequent high-stakes math test than comparable students at schools with no failing racial subgroups. Estimates of this impact are of similar magnitude to the test score decrease that occurs when students switch schools midyear and occur after controlling for individual student pre-test low-stakes math scores and general differences arising between failing and successful schools. Consistent with the strategic instruction hypothesis, this difference increases as failing schools face more severe NCLB sanctions and in schools that *ex ante* are more likely to fail to make AYP. These impacts occur despite controlling for the racial and socio-economic makeup of schools and a host of individual student characteristics. These findings are also robust to controlling for non-random sample attrition and do not appear to be the result of administrators targeting students by their *a priori* beliefs of student ability.

A handful of researchers have investigated a form of strategic instruction based not on race but on student ability. Chakrabarti (2007) uses school-level data to analyze the behavior response of schools threatened under Florida's "opportunity scholarship" program. This program predates NCLB but 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 for administrators to focus on students who are marginally below the threshold required to pass Florida's high-stakes test. When compared to students at similar but non-threatened 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.

Using student-level data, Reback (2008), Neal and Schanzenbach (2007) and Springer (2008) document the presence of strategic instruction based upon a student's expected ability to influence a school's AYP determination. Reback compares students within buildings and 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 impact their school's ratings. Applying a difference-in-difference estimator to Chicago Public School data, Neal and Schanzenbach find that students in the middle of the distribution on a low-stakes test achieve significantly higher gains on a high-stakes test after the adoption of each of two different accountability systems. Likewise, Springer shows that low achieving students at schools under NCLB sanction threat gain more than comparable

[2] p.11.
[3] p. 117.

students at a non-threatened school. Moreover, Springer finds that the gains of the low achieving students do not come at the expense of high achievers and suggests that the NCLB causes threatened schools to reduce their "operational slack" to make these improvements.

Before proceeding, one caveat is necessary. The presence of strategic instruction may not result in an inefficient outcome. If, prior to NCLB, schools over-expended resources on students of would-be successful racial groups, then NCLB incentives discussed here may improve overall resource allocation. Further, as suggested by Springer (2008) and Chakrabarti (2007), if building administrators respond to NCLB by introducing more effective teaching techniques, better curriculum, or a more efficient use of resources, then NCLB may improve overall student learning and these improvements may be significantly larger than any redistribution that occurs between subgroups.

2. NCLB and student testing in Washington

As part of an initiative to increase educational accountability, Washington introduced the Washington Assessment of Student Learning (WASL), a statewide test of reading, writing, listening, and mathematics in 1997.[4] The WASL became the high-stakes test used to identify AYP when NCLB was enacted. The 4th grade WASL tests mathematics, reading, listening and writing. In order to avoid complications that arise when combining scores from tests of different subjects, this paper analyzes only the WASL math results which have been normalized to mean zero and variance equal to one within each year.

NCLB requires school districts to bring all students to the "proficient" level in reading and mathematics by 2013–2014.[5] In the meantime, individual schools must meet state AYP targets towards this goal for both their overall student population as well as for eight socio-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 the WASL and compares this to the state-imposed pass rate.[6] 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.[7] However, this condition holds only if the subgroup under consideration contains at least thirty students. For example, a school with less than thirty limited English students must be above the required pass rate for the other eight categories but is automatically deemed acceptable in the limited English category. As designed, a single student can be a member of many groups and therefore impact 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 of percent proficient as well as the calculation of the three subgroups. In 2005 (the final cohort examined in this paper), 62.7% of 4th graders demonstrated proficiency in both the reading and the math sections of the WASL and 64.7% demonstrated proficiency on the math portion. Indeed, the vast majority of schools failing to make AYP in 2005 did so because of a failure to achieve the required pass rate in mathematics. Of the 207 Washington buildings failing to make AYP that year, 161 were due to poor math scores.

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 academic level. The ITBS is given in Washington near the end of the student's 3rd grade year, the year immediately prior to the WASL. Using the ITBS math results presents a number of advantages. First, since the ITBS is not employed as a tool to determine AYP, it is unlikely to be the direct focus of strategic instruction. Instead it may be a tool used by administrators who decide how to allocate future resources across students. Secondly, since the ITBS is given the year previous to the WASL, it can measure student ability. Another advantage conveyed with the ITBS data is the large number of demographic, social-economic, and academic variables observed. These variables are used as explanatory variables in later regressions. Unlike the WASL scores, a student's ITBS is measured as a percentile relative to all nationwide 3rd graders taking the ITBS.

Optimally, a researcher would compare schools under NCLB with those that were not impacted by NCLB to test if strategic instruction took place. But, since 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 prac-

[4] Much of this section describes the Washington testing system as it was in place during the time which this paper addresses. Since that time, Washington has made a number of changes to its high-stakes testing system including replacing the WASL with a online exam. Washington also altered the method it used to determine AYP but these changes postdated the data used in this paper.

[5] The raw math WASL score required to meet proficiency is 400. After normalizing to mean zero and standard deviation of one, this translated into −.19 standard deviations in 2003–2004 and −.25 standard deviations in 2004–2005.

[6] During the period covered by this data, required pass rates in Washington were calculated by determining the average annual improvement needed between 2001 and 2002, when NCLB was implemented, and 2013–2014 when 100% of students are required to demonstrate proficiency. For example, in 2001–2002, 29.7% of 4th grade students were rated as math proficient by Washington. 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 2013–2014. Thus, the mathematics required pass rate required to make AYP in the 2002–2003 school year was 29.7% + 5.86% = 35.56%. Finally, AYP is granted only if 95% of all continuously enrolled students at each grade level take the WASL.

[7] The actual AYP calculations also include a confidence interval adjustment in the schools favor that depends upon the number of test takers in a given year. This adjustment typically adds two or three percentage points to the actual percentage of students demonstrating proficiency in a given year. Also, AYP is awarded to schools that make "safe harbor." These schools are those that, while containing too few proficient students to be above the state's pass rate, did make significant progress over the prior year that they are awarded AYP. In 2005, 11 Washington schools that included a fourth grade made AYP through the safe harbor provision.

tices. However, as suggested by Rouse et al. (2007), 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 the WASL have been given in Washington since the mid-1990s, which creates the possibility of a difference-in-difference identification strategy. If NCLB creates strategic instructional behavior, then differential WASL outcomes should be found among schools under the threat of sanctions and should be present only after NCLB was enacted.

3. Data and descriptive statistics

The data used in this article is the same as used in Krieg (2008) and is more fully described therein. This data consists of four cohorts of paired observations of ITBS/WASL scores for third/fourth graders. The first observed cohort of third 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 schools 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 their building administrator knew their building's AYP status. Students in the final two cohorts began their WASL year after schools knew their AYP status so administrators had greater ability to pursue strategic instruction for these students. The heterogeneity between these two sets of cohorts offers one opportunity for identifying the impact of NCLB.

As a first attempt to investigate strategic instruction, the final two cohorts are examined—the cohorts who took the WASL in buildings where AYP status was known prior to the beginning of the school year. After excluding special education students and those with missing observations, the pooled number of student observations in the last two cohorts is 112,485. This represents 74.8 percent of all Washington public 4th grade students and 85.7 percent of all non-special education students. Panel A of Table 1 divides this cohort into two groups: students at schools who made AYP in the previous year and students who are members of a race that made AYP in the previous year but who attend buildings which did not make AYP because another racial group failed. For example, this second group includes Hispanic students if Hispanics at their building made AYP but Whites did not. In this example, Panel A of Table 1 would not include those White students.

Panel A of Table 1 demonstrates a number of important features. First, a significant difference in WASL and ITBS performance occurs between passing students at AYP schools and passing students at failing schools. On average, students at AYP schools score .085 standard deviations above the state WASL mean and average just above the 62nd percentile on the ITBS. Members of a successful racial group at a failing school score .27 standard deviations below the average on the WASL and are at the 56th ITBS percentile. Since Table 1 includes only those students in racial groups that made AYP, this difference is not caused by inclusion of failing groups of students. Rather, this difference likely arises from any number of factors. For instance, AYP schools have roughly half the free/reduced lunch population relative to non-AYP schools and students of successful racial groups at non-AYP schools are much more likely to be minorities than those at AYP schools. These factors will be controlled for in a more complete econometric model.

If strategic instruction occurs, passing students at failing schools will make smaller gains between their ITBS test year and their WASL year than do observationally equivalent students at AYP schools. Panel A of Table 1 suggests that this may be the case. Students at failing schools averaged at the 56th percentile on the 3rd grade ITBS. Those at passing schools averaged at the 62nd percentile, a statistically significant but relatively small difference in performance. In their fourth grade year, students at failing schools averaged .27 standard deviations below the WASL average (the 34th WASL percentile), those at passing schools averaged almost .1 standard deviation above average (the 49th WASL percentile). Consistent with the strategic instruction hypothesis, the larger difference in WASL performance relative to ITBS performance suggests significant academic improvement of students at passing schools relative to students in passing subgroups at schools that failed. Of course these differences could be explained by a number of competing hypotheses other than strategic instruction. For instance, the difference between a 56th and 62nd ITBS percentile student may be large in terms of academic competency making it difficult for schools to transform low ITBS students into high WASL scorers. To help distinguish between these possibilities, consider Panel B of Table 1. Panel B presents descriptive statistics for the two cohorts that were observed prior to AYP status being defined in Washington. These two cohorts are separated into two groups: students at buildings that will make AYP in the future and students of a racial group that will make AYP in the future who attend schools who will not make AYP. Thus, Panel B simply presents the same schools and racial groups as does Panel A but does so for the two years prior to NCLB.

Contrasting students at AYP schools before and after the imposition of NCLB (the first columns of Panels A and B) demonstrates little difference in either ITBS or WASL scores. Of course, little difference is to be expected in performance if those schools expect to make AYP and do not alter their instructional practices as a result of NCLB. However, consistent with the strategic instruction hypothesis, the relative performance of students of a passing race at a failing schools appears higher prior to NCLB. These students averaged at the 54th percentile on the pre-NCLB ITBS, a small difference from their post-NCLB average of the 56th percentile. However, these students scored .21 standard deviations below the pre-NCLB WASL average (the 44th WASL percentile); a rather large improvement over the −.27 standard deviation (the 34th percentile) post-NCLB WASL average. This decrease of relative performance after the NCLB is explored systematically in the next section.

4. Econometric evidence

The preceding descriptive statistics suggest that the gains made for students in a passing racial group at a failing

Table 1
Descriptive statistics.

Variable	Panel A: cohorts where AYP is known		Panel B: cohorts where AYP is unknown	
	Students at non-failing schools	Students of race other than that which failed	Students at future non-failing schools	Students of race other than that which will fail
Student variables				
WASL	0.085 (0.977)	−0.270 (0.958)	0.096 (0.963)	−0.209 (0.956)
ITBS	62.70 (28.40)	56.06 (28.76)	61.63 (28.49)	54.11 (28.75)
Indian	0.024 (0.153)	0.026 (0.159)	0.023 (0.151)	0.094 (0.292)
Asian	0.080 (0.271)	0.080 (0.271)	0.075 (0.264)	0.079 (0.270)
Black	0.052 (0.221)	0.057 (0.232)	0.052 (0.222)	0.049 (0.216)
Hispanic	0.130 (0.336)	0.114 (0.318)	0.108 (0.311)	0.092 (0.289)
White	0.707 (0.455)	0.717 (0.450)	0.730 (0.444)	0.685 (0.464)
English never	0.123 (0.328)	0.128 (0.334)	0.110 (0.313)	0.122 (0.327)
Building variables				
% Indian	2.55 (5.78)	4.66 (12.76)	2.47 (5.00)	8.96 (18.96)
% Asian	8.15 (8.74)	4.31 (7.14)	7.69 (8.21)	4.20 (6.96)
% Black	5.57 (8.25)	7.15 (14.31)	5.69 (8.60)	7.12 (14.00)
% Hispanic	13.76 (18.05)	51.51 (24.83)	11.85 (16.45)	45.24 (26.71)
% Free/reduced	39.35 (22.97)	74.17 (15.65)	37.47 (22.70)	72.01 (14.81)
Enrollment per grade	76.40 (26.87)	86.41 (23.15)	76.97 (27.59)	82.74 (19.84)
Number of obs.	110,755	1730	113,933	1820
Number of schools	287	25	283	25

school were larger pre-NCLB than post-NCLB while those at successful schools demonstrate no such difference. To explore this further, consider the regression:

$$\begin{aligned} \mathrm{WASL}_{itb} = {} & \beta_0 + \sum_{j=1}^{L} \phi_j \mathrm{ITBS}_{itb}^{j} + \alpha \mathrm{AYPFAIL}_{tb} + \gamma \mathrm{AYPFAIL}_{tb} \\ & * \mathrm{IN\_PASSING\_SUBGROUP}_{itb} + \lambda \mathrm{NCLB}_t + \boldsymbol{\psi} \mathbf{B}_{tb} \\ & + \boldsymbol{\lambda} \mathbf{X}_{itb} + \varepsilon_{itb} \end{aligned} \quad (1)$$

where WASL_{itb} is student i's test score during time period t in building b. $\mathbf{X}_{itb}$ is a matrix of student-specific control variables, and $\mathbf{B}_{bt}$ represent a matrix of time-varying building control variables.[8] AYPFAIL is a binary variable equaling one if the building failed to make AYP in the previous year. Since no buildings failed to make AYP during the first two observed years, AYPFAIL is equal to zero for all of these observations. NCLB is a binary variable equaling one for the last two observed cohorts (those cohorts who took the WASL after full NCLB implementation). Because of the non-linear relationship between tests measured on a percentile basis and those measured in standard deviations around the mean, WASL test scores are assumed to be a polynomial function of ITBS scores where the degree of polynomial, *L*, is determined by minimizing the AIC.

The variable of interest in Eq. (1) is AYPFAIL*IN_PASSING_SUBGROUP. If a student is a member of a subgroup which demonstrated proficiency in the prior year, they are assigned a value of 1 for IN_PASSING_SUBGROUP; otherwise they are assigned zero. Therefore, the interactive variable AYPFAIL*IN_PASSING_SUBGROUP equals one for students of a racial group which made AYP in the previous year and who attend a school that failed to make AYP because of the failure of another racial group. Because of the inclusion of the variables AYP and NCLB, AYPFAIL*IN_PASSING_SUBGROUP measures the marginal impact on successful subgroups relative to the overall impact of failing to make AYP and relative to pre-NCLB expected performance. If strategic instruction occurs, then γ, the coefficient on AYPFAIL*IN_PASSING_SUBGROUP, would be less than zero indicating that WASL performance was lower for students of a successful racial group who attend schools that failed because of the prior performance of another racial group.

Panel A of Table 2 presents estimates of Eq. (1). Students of races that did not cause the previous AYP failure expect to score .050 standard deviations lower on the WASL than similarly situated students at passing schools. To put this into context, the (unreported) coefficient on black (relative to white) is −.146, the impact of changing schools midyear is −.048, of having been held back at least one grade is −.092, and of having a computer at home .060. Thus, a student of a passing racial group who attends a school that had another racial group fail is expected to have their WASL performance diminish by about the same amount as would occur as if he or she changed schools midyear or about the same as the difference that occurs between students with and without a computer at home.

Of interest in Panel A of Table 2 is the negative coefficient associated with failing to make AYP. All students at schools who failed to make AYP in the prior year can be expected to score .044 WASL standard deviations lower than schools making AYP in the prior year. Among other things, this

[8] The student control variables include nine binaries representing ethnicity, five binaries representing the duration of student enrollment in the school, four binaries indicating their frequency of reading for fun, six binaries indicating their amount of daily television watched, three binaries indicating the frequency of speaking English at home, gender, the amount of computer usage at school, the presence of a computer at home, and if they skipped or were held back a grade. The building control variables include the percentage of student body in each of the five NCLB racial groups, the percent of students receiving free or reduced lunches, the average building enrollment and its square, and five binary variables indicating the building type (traditional elementary, comprehensive, parent partnership program, internet/computer school, or alternative school).

Please cite this article in press as: Krieg, J.M. Which students are left behind? The racial impacts of the No Child Left Behind Act. *Economics of Education Review* (2011), doi:10.1016/j.econedurev.2011.02.004

Table 2
OLS estimates of WASL scores.

	Variable	Panel A	Panel B	Panel C	Panel D
λ	NCLB	.030*** (.003)	.031*** (.003)	.032*** (.004)	.015*** (.004)
α	AYPFAIL	−.044*** (.008)	−.031*** (.008)	−.049*** (.008)	−.039*** (.008)
υ	AYPFAILTWICE		−.090*** (.021)		
γ	AYPFAIL*IN_PASSING_SUBGROUP	−.050** (.021)	−.054** (.022)	−.049*** (.021)	−.055*** (.020)
ς	AYPFAILTWICE*IN_PASSING_SUBGROUP		−.035* (.019)		
	ENGLISH AT RISK				−.014** (.006)
	LOW INCOME 30				−.021*** (.004)
	R^2	.577	.577	.584	.577
	N	228,238	228,238	170,272	228,238
	L	10	10	8	10
	F-test of $\gamma=\psi=0$		4.27**		

Notes: Standard errors corrected for clustering within buildings are in parenthesis. All regressions contain the independent variables listed in note 8.
* Statistical significance at the 10% level.
** Statistical significance at the 5% level.
*** Statistical significance at the 1% level.

may be the result of unobserved differences in student composition, teacher recruitment and retention, and financial constraints between passing and failing buildings. It is important to note that the strategic instruction finding occurs in the presence of this AYP status; in other words students of a successful subgroup at a failing school expect to score .094 (=.044 + .050) standard deviations worse than students at a passing school.

Panel A of Table 2 also presents an estimated coefficient associated with the variable NCLB. Conditional expected WASL tests scores are .03 standard deviations higher after enactment of NCLB—about 30% of what students of successful racial groups at failing schools expect to lose relative to their counterparts at successful schools. Potential explanations for this improvement are many: NCLB better focused resources on academics, teachers may teach to the test, or the resources associated with NCLB were used efficiently by school administrators. Regardless of the cause, Table 2 presents a picture of a change in relative racial performance among students at failing schools and a simultaneous small, but statistically significant, increase in overall test performance.

If the estimate of $\gamma=-.050$ is the result of administrators focusing attention on racial groups who previously failed to make AYP, then this difference should grow in magnitude at schools that have failed to make AYP in consecutive years resulting in more severe NCLB sanctions. In this data, it is possible to identify schools and races that have failed AYP for two consecutive years. Consider the regression:

$$\begin{aligned}\text{WASL}_{\text{itb}} = \beta_0 + \sum_{j=1}^{L}\phi_j \text{ITBS}^{j}_{\text{itb}} + \alpha\text{AYPFAIL}_{\text{tb}} \\ + \nu\text{AYPFAILTWICE}_{\text{tb}} + \gamma\text{AYPFAIL}_{\text{tb}} \\ *\text{IN\_PASSING\_SUBGROUP}_{\text{itb}} + \zeta\text{AYPFAILTWICE}_{\text{tb}} \\ *\text{IN\_PASSING\_SUBGROUP}_{\text{itb}} + \lambda\text{NCLB}_t + \boldsymbol{\psi}\mathbf{B}_{\text{bt}} \\ + \boldsymbol{\lambda}\mathbf{X}_{\text{itb}} + \varepsilon_{\text{ibt}}\end{aligned} \tag{2}$$

where AYPFAILTWICE equals one if racial groups at the building failed for two consecutive years. AYPFAILTWICE*IN_PASSING_SUBGROUP equals one if a student is a member of a racial group who successfully made AYP for two consecutive years and attends a school where another racial group failed to make AYP for two consecutive years.[9] If strategic instruction occurs, one would expect ζ to be negative suggesting a further decrease in academic performance.

Panel B of Table 2 presents estimates of selected coefficients from Eq. (2). These results support the strategic instruction hypothesis. Students of a passing racial group at a school failing in the previous year expect to score .055 standard deviations lower on the WASL than similar students at passing schools. However, students of a passing racial group at a school that failed for two consecutive years can be expected to lose an additional .035 standard deviations. This larger decline would be expected if school administrators focused increased attention to needy groups of students at the expense of those that have traditionally performed adequately on the WASL.

5. Robustness checks

Tables 1 and 2 suggest that students of a successful racial group at schools that fail to make AYP perform worse than similarly situated students at passing schools. While this may be due to strategic instruction, alternative hypotheses are possible and explored in this section.

One possible concern is that building administrators predicted the passage of NCLB and initiated strategic instruction prior to the enactment of the legislation. Since Eq. (1) relies upon comparing student performance before and after NCLB, an administrator acting in such a way would obscure the differences found in a before and after estimation technique. To check for this possibility, Eq. (1) is re-estimated without data from the second cohort which is the group of students who began their high-stakes test year prior to enactment of NCLB but concluded the year (and took the WASL) after the legislation passed. If administrators were able to impact a student's learning during this

[9] In the case of consecutive failure, students are assigned the value of one to each of AYPFAILURERACE and AYPFAILURERACETWICE. Buildings are assigned the value of one to each of AYPFAIL and AYPFAILTWICE.

year, it is this group of students who would be impacted. Panel C of Table 2 presents estimates of Eq. (1) without this cohort resulting in a comparison of the final two cohorts with the first cohort. When the second cohort is removed the estimated impact of attending a school with a failing subgroup is nearly identical to the full sample results (Panel A).

A second concern involves the fact that any particular student may be a member of more than one NCLB group that contributes to AYP determination. Given the fact that special education students are removed from this analysis, at most a particular student may be a member of two non-racial NCLB groups: low-income and limited English. Because of non-random sorting of race across these two groups, it is possible that failing to account for membership of these groups biases the AYPFAIL*IN_PASSING_SUBGROUP estimates. For instance, it is possible that a student is a member of a successful racial group but is also a low-income student at a school that failed to make AYP because of low performance by low-income students. This particular student might receive increased academic attention from teachers and administrators and, in turn, perform better on the WASL than expected. If this were the case, the AYPFAIL*IN_PASSING_SUBGROUP estimates would be upwardly biased.

To account for the possibility of being a member of a low-income or limited English group, two additional variables are added to Eq. (1). The first, ENGLISH AT RISK, is a binary variable equal to one if a student meets two conditions: (1) they do not speak only English at home and (2) their NCLB cohort contains at least 30 limited English students.[10] The second condition is a prerequisite for the NCLB to apply to that subgroup for the school. Presumably, a school with less than 30 limited English students will spend little effort strategically instructing based upon that category. The second variable, LOW INCOME 30 is a binary variable equal to one if the student's NCLB cohort contains thirty or more low income students.[11] Both LOW INCOME 30 and ENGLISH AT RISK control for the possibility that an individual student receives strategic instruction because of their membership in multiple NCLB subgroups. Panel D of Table 2 presents estimates of Eq. (1) augmented by LOW INCOME 30 and ENGLISH AT RISK. Relative to Panel A, the inclusion of LOW INCOME 30 and ENGLISH AT RISK causes the AYPFAIL*IN_PASSING_SUBGROUP to fall by about only ten percent suggesting earlier estimates of this coefficient were minimally biased upwards. In other words, it appears members of a successful racial group who attend a failing school do worse than their peers at passing schools even after controlling for the possibility of being members of other NCLB subgroups.

While strategically targeting students based upon their race may be plausible, school administrators are privy to information that could result in a different form of strategic instruction that this empirical strategy mistakenly identifies as being based upon race. As suggested by Reback (2008), student test history presents administrators with rough estimates of each student's propensity to show proficiency on a high-stakes test. Rather than targeting students based upon race, an administrator could target instructional resources using ITBS test history and their perceptions of individual student ability. For instance, based upon their 3rd grade ITBS score, an administrator could place students on the perceived margin of passing the WASL in a strong 4th grade teacher's classroom and place very strong and very weak ITBS students with less able teachers.[12] This would increase learning for middle-ability students and cause lower gains for students on the tails of the ability distribution. As long as high ability students continue to pass the WASL, this strategy would maximize the percent of students passing the WASL and therefore the school's probability of making AYP. If administrators behave this way and if test scores are correlated with race, then the results from Tables 1 and 2 would occur not because of strategic instruction focusing on student race, but rather because students receiving decreased attention are those whose previous tests scores are perceived by administrators as being those requiring the least academic attention.

One way of testing for this possibility is to interact AYPFAIL with each students' 3rd grade ITBS score, an identification strategy used by Krieg (2008). If administrators at schools that failed to make AYP in the previous year direct resources away from students who scored well on the 3rd grade test, then this interacted variable (AYPFAIL × ITBS) will be negative and its presence should cause the significant coefficients associated with AYPFAIL*IN_PASSING_SUBGROUP to become insignificant. To control for non-linearities in the relationship between the ITBS and WASL, polynomials of ITBS interacted with AYPFAIL and are included in:

$$\begin{aligned}\text{WASL}_{itb} = {} & \beta_0 + \sum_{j=1}^{L}\phi_j \text{ITBS}_{itb}^{j} + \sum_{j=1}^{M}\xi_j \text{AYPFAIL}_{tb} \times \text{ITBS}_{itb}^{j} \\ & + \alpha\text{AYPFAIL}_{tb} + \gamma\text{AYPFAIL}_{tb} \\ & * \text{IN\_PASSING\_SUBGROUP}_{itb} \\ & + \lambda\text{NCLB}_t + \boldsymbol{\psi}\mathbf{B}_{bt} + \boldsymbol{\lambda}\mathbf{X}_{itb} + \varepsilon_{ibt} \qquad (3)\end{aligned}$$

where L and M and determined by minimizing the AIC.

Table 3 presents estimates of ξ which are individually insignificantly different from zero (though they do jointly explain WASL results). The coefficient on AYPFAIL*IN_PASSING_SUBGROUP remains negative, statistically significant, and of slightly larger magnitude than the

[10] The State of Washington does not release data that identifies which individual students are members of the limited English subgroup. However, the ITBS test asks each respondent if they speak only English at home, some English at home, and never speak English at home. The strategy used here assumes that each respondent claiming they either speak some English or never speak English at home are in the limited English subgroup.

[11] Optimally, one would want to know if a particular student is low income but the State of Washington does not release this information nor is it identified on the ITBS survey.

[12] This form of strategic instruction was explored by Krieg (2008), Reback (2008), and Chakrabarti (2007).

Table 3
OLS estimates of WASL scores using interacted student ITBS scores.

λ	NCLB	.030*** (.003)
α	AYPFAIL	−.114*** (.032)
γ	AYPFAIL*IN_PASSING_SUBGROUP	−.058*** (.021)
ξ_1	AYPFAIL × ITBS	.0007 (.002)
ξ_2	AYPFAIL × ITBS^2	.00004 (.00005)
ξ_3	AYPFAIL × ITBS^3	−.0000005 (.0000004)
	R^2	.577
	N	228,238
	L	10
	M	3
	F-test of $\xi_1 = \xi_2 = \xi_3 = 0$	8.64 (.000)

Notes: Standard errors corrected for clustering within buildings are in parenthesis. All regressions contain the independent variables listed in note 8.
*** Statistical significance at the 1% level.

OLS estimates of Eq. (1). This suggests that strategic instruction findings are not solely based upon prior test scores but also upon subgroup membership.

Another robustness check involves sorting the sample to control for school outliers. The possibility exists that, based upon the composition of their student bodies, some schools are so certain of making AYP (or so certain of failing to make AYP) that administrators face no incentive to perform strategic instruction. If this is the case, then the prior results may understate the impact of strategic instruction in those schools that perform it.

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

$$PR(Y_b = 1) = f(\boldsymbol{\psi}\mathbf{B}_b + \varepsilon_b) \quad (4)$$

where Y is equal to zero if a building failed to make AYP in either of 2004 or 2005 (the last two observed cohorts), and **B** represents the building control variables used in Eqs. (1), (2) and (3) measured in 2001. This logit can be thought of as a forecast of which schools will make AYP based upon their characteristics observed at the time of NCLB enactment. From this logit, predicted probabilities of a building making AYP are generated, sorted, and divided into the lowest, middle, and highest thirds. Using the entire sample of students, Eq. (1) is then re-estimated for each third and results are presented in Table 4.

Table 4 presents evidence that all schools act strategically with respect to student race, but the predominant effects occur at schools in the lower third of the predicted probability of making AYP. Students of a passing race at schools in this group who failed to make AYP can expect to score .062 WASL standard deviations lower than similar students at passing schools in this group. Schools in the middle and top thirds of the probability of making AYP have less evidence of strategic instruction. In both cases, passing races expect to score .021 standard deviations worse than comparable students however neither of these measures differ statistically from zero. This pattern of findings is consistent with administrators at schools perceived to be in danger of failing to make AYP acting aggressively by redirecting resources towards racial groups that may cause the failure. Schools less likely to fail have much less urgency in following this course of action and a much smaller racial impact results.

A final explanation for these findings is that the composition of students taking the WASL differs between AYP and non-AYP schools and this difference is not accounted for by the independent variables in the regressions. This concern has been addressed by a number of studies, especially with regard to strategic placement of students in special education programs[13] and through strategic administrative exclusion of students most likely to fail their high stakes test.[14] If non-random selection of students omitted from this analysis occurs, then the results of the prior regressions may be biased in favor of finding strategic instruction.

Panel A of Table 5 presents counts of included and missing observations of general education students by year. Over the time period examined, the percentage of valid general education students with complete WASL and ITBS observations has remained stable suggesting that NCLB did little to change the trend of missing exams. Secondly, the numbers of missing observations are relatively small; over the four cohorts observed less than 15% of all Washington general education students are missing. Unless there is a high correlation between being unobserved and WASL performance, this small number of missing observations is unlikely to overturn the prior results. Panel B of the same table explores this possibility by comparing the missing observation percentage of the treatment group (those AYPFAIL*IN_PASSING_SUBGROUP = 1 for the last two cohorts and, for the first two cohorts, members of those subgroups who will pass but attend failing schools) with the control group. As seen in Panel B, about 16.4% of potential treatment students are excluded from the analysis because of missing test scores; a rate about one-sixth more frequent than students in the control group. Like the overall percent of missing observations, excluded rates are very stable over the four cohorts and between control and treatment groups. While the missing observation rate is stable over time, one can imagine failing schools encouraging some students to take the WASL while simultaneously discouraging others in hopes of making AYP. This non-random attrition needs exploration before making the conclusion that strategic instruction exists.

To test for the possibility of sample selection bias, a two-stage Heckit procedure is employed.[15] In the first stage, a probit augments the regressors from Eq. (1) with the contemporaneous percentage change of a county's population to estimate if a student missed the WASL. Because a possible reason for missing the WASL is that students move from their local school district, including the percentage change in the local population theoretically explains sample attrition. The second stage of the Heckit procedure adds the inverse Mills ratio from this probit to Eq. (1). Results from these two regressions are presented in Panel A of Table 6.

Analysis of the first stage probit in Panel A of Table 6 reveals that students do miss the WASL systematically. Schools in counties with high population growth are more

[13] See Figlio and Getzler (2002), Deere and Strayer (2001), Cullen and Reback (2006), and Jacob (2005).
[14] Figlio (2006) finds that during test weeks in Florida, the duration and frequency of disciplinary suspensions for low-performing students in grades that face high-stakes tests increases.
[15] See Wooldridge (2002, chap. 17) for details.

Table 4
OLS estimates of WASL scores by building likelihood of making AYP.

		Lowest third	Middle third	Highest third
λ	NCLB	.030*** (.006)	.031*** (.005)	.037*** (.005)
α	AYPFAIL	−.052*** (.010)	−.034 (.023)	−.065*** (.023)
γ	AYPFAIL*IN_PASSING_SUBGROUP	−.062*** (.021)	−.021 (.061)	−.021 (.073)
	R^2	.563	.562	.552
	N	68,930	73,880	85,428

Notes: Standard errors corrected for clustering within buildings are in parenthesis. All regressions contain the independent variables listed in note 8.
*** Statistical significance at the 1% level.

Table 5
Missing observations.

Panel A: numbers and percentages of excluded students

Academic year	General education	Missing ITBS score	Missing WASL score	Missing both ITBS and WASL	Valid observations
2001–2002	67,346	4922 (7.3%)	3808 (5.6%)	829 (1.2%)	57,787 (85.8%)
2002–2003	67,878	5323 (7.8%)	3944 (5.8%)	645 (1.0%)	57,966 (85.4%)
2003–2004	65,583	4756 (7.3%)	3500 (5.3%)	667 (1.0%)	56,660 (86.4%)
2004–2005	65,669	5124 (7.8%)	3855 (5.9%)	865 (1.3%)	55,825 (85.0%)
Total	266,476	20,125 (7.5%)	15,107 (5.7%)	3006 (1.1%)	228,238 (85.6%)

Panel B: percentages of missing observation by control and treatment group

Academic year	Percent of treatment group missing either ITBS or WASL	Percent of control group missing either ITBS or WASL
2001–2002	16.4%	13.9%
2002–2003	16.5%	14.4%
2003–2004	16.1%	13.4%
2004–2005	16.0%	14.8%
Total	16.4%	14.1%

Notes: The treatment group consists of two groups: (1) for the final two cohorts, students of a passing subgroup at a failing school and (2) for the first two cohorts, students who are members of a subgroup who will pass in the future at a school that will fail in the future.

Table 6
Estimates of WASL scores controlling for non-random attrition of students.

		Panel A		Panel B	Panel C	
		1st Stage probit	2nd Stage Heckit	OLS with estimated ITBS	1st Stage probit	2nd Stage Heckit
λ	NCLB	.012** (.006)	.028*** (.003)	.031*** (.003)	.010* (.006)	.026*** (.003)
α	AYPFAIL	−.209*** (.019)	.003 (.012)	−.040*** (.006)	−.202*** (.018)	.077*** (.007)
γ	AYPFAIL*IN_PASSING_SUBGROUP	.090*** (.037)	−.087** (.041)	−.060*** (.017)	.088*** (.037)	−0.071*** (.027)
	Population growth rate	.037*** (.004)	–	–	.014*** (.004)	
	Inverse mills ratio	–	.911*** (.178)	–	–	3.28*** (.094)
	R^2		.575	.628		.632
	N	243,345	243,345	248,363	263,470	263,470

Notes: The dependent variable for the 1st stage probit equals one if the observation did not take the WASL and equals zero otherwise. Standard errors corrected for clustering within buildings are in parenthesis. All regressions contain the independent variables listed in note 8.
* Statistical significance at the 10% level.
** Statistical significance at the 5% level.
*** Statistical significance at the 1% level.

likely to enroll students who later miss the WASL. After the passage of NCLB, the conditional probability of individuals missing the WASL declined. Further, students at schools which failed to make AYP in the previous year are also less likely to miss the WASL. This may be an artifact of failing schools more aggressively recruiting additional test takers in hopes of improving past performance. Finally, students of a passing racial group at a school that failed to make AYP are more likely to miss the WASL. Possibly, these students do not receive the encouragement to take the WASL from their administrators to the same extent as those in failing racial groups. Whatever the reason, if these students are stronger than average test takers, the prior regression results would actually understate the impacts of strategic instruction.

The second stage Heckit results in Panel A of Table 6 explore the possible biases that occur because of WASL attrition. Relative to the estimates of Eq. (1), the Heckit results suggest that the impacts of racial strategic instruction are actually larger after correcting for non-random WASL attrition. Those students of a successful racial group are expected to score .087 standard deviations worse than comparable students, a large, but statistically insignificant difference when compared to the OLS estimates.[16]

A related non-random attrition concern exists. If high ability students are more likely to leave a school that recently failed to make AYP, then estimates of student performance of remaining students may appear lower leading to the strategic instruction conclusion. To check for this possibility, the interactions of AYPFAIL and polynomials of student ITBS scores were included in the first stage probit of Panel A, Table 6. If high ability students are more likely to leave failing schools than lower ability students, coefficients on these variables will be positive. It turns out, these coefficients are individually and jointly insignificant ($F = .1923$, $p = .901$) suggesting that non-random attrition by ability and school NCLB status does not occur.

Another possible selection bias is suggested by Table 5. A significant number of 3rd graders miss the ITBS test. Since this test is an explanatory variable of all prior regressions, its omission may bias estimates of the WASL results if students miss the ITBS in a non-random fashion. However, since the ITBS is an explanatory variable, the Heckit procedure cannot be used to explore if omitting those who failed to take the ITBS biases the regression coefficients. To check for the importance of missing the ITBS, a second two-stage procedure is followed. The first stage employs the subsample with complete observations of ITBS scores and estimates the regression:

$$\text{ITBS}_{\text{ibt}} = \lambda \mathbf{X}_{\text{ibt}} + \boldsymbol{\psi}\mathbf{B}_{\text{bt}} + \sum_{j=1}^{m} \theta_j \text{WASL}_{\text{ibt}}^{j} + \tau t + \varepsilon_{\text{ibt}} \qquad (5)$$

where **X** and **B** are defined as in Eq. (1). Using the estimated coefficients from Eq. (5), predicted ITBS scores are generated only for those students with missing ITBS scores. The students with generated ITBS scores are then integrated into the sample and Eq. (1) is re-estimated with results presented in Panel B of Table 5. Results from this procedure are broadly consistent with those in Table 2; WASL scores are .060 standard deviations lower for students of passing racial groups at failing schools and the impact of NCLB and school AYP failure are similar as the prior OLS estimates.

As a final check for sample selection bias, the preceding two analyses were merged. Using Eq. (5), ITBS scores were created for those individuals missing the ITBS and then integrated into the complete data set. Using these data, a second Heckit accounting for missing WASL scores was estimated and results presented in Panel C of Table 5. The results of this were broadly similar with those of the prior Heckit model. Students of a passing race at a failing school are expected to score .071 standard deviations lower on the WASL than comparable students. Taken as a whole, it does not appear that non-random WASL nor ITBS attrition accounts for the subgroup impacts of NCLB.

6. Discussion and conclusions

This article demonstrates a differential impact of NCLB on racial groups depending upon their and other racial groups' prior success on a high stakes test. Students of a successful racial group at a school where another racial group failed to make AYP are expected to score .050 standard deviations lower on Washington's high stakes test than are similar students who attend a school where no racial group failed. This test difference is of similar magnitude to the conditional impact of switching schools midyear and the conditional differences occurring between students having and not having computers at home. This finding occurs in the presence of individual controls for prior standardized test scores, demographic features, and individual student characteristics. It also occurs in the presence of building level controls for prior AYP passage, racial make-up, enrolment, building type, and the percent of students receiving free/reduced lunch. The estimated impact of this disparity grows in magnitude as the building fails to make AYP in consecutive years and faces more significant NCLB sanctions. This finding is also stronger at schools that are *a priori* more likely to fail to make AYP. Further, this occurs even after controlling for a second type of strategic instruction that may occur when building administrators target resources towards students based upon prior test scores. Finally, these findings are robust to non-random sample attrition from the WASL and ITBS tests. Taken as a whole, this evidence suggests that building administrators participate in strategic instruction; that is, administrators focus their efforts on racial groups that have trouble making AYP. Given the limits on school resources, this redirection of resources towards one racial group causes a diminution in academic performance of students in successful racial groups.

Two arguments may be made that these findings underestimate the true impact of NCLB. First, consider a school where each racial group made AYP in the prior year but one group was close to failure. Because the required pass rate rises each year, should this school fail to increase the performance of the group that barely passed, it will fail in future years. A school in this position has incentives to perform strategic instruction prior to failure but, under the empirical strategy used in this paper, would not be identified as doing so. Thus, the estimated impacts of strategic instruction may understate the actual impacts of inter-subgroup resource shifting. The second issue has to do with the minimum size of the racial group required to determine AYP failure. Under NCLB, schools with fewer than 30 students in a demographic group automatically receive AYP for that group. This requirement reduces the incentive to perform strategic instruction at small schools however, schools that are close to this limit may participate in strategic instruction because an unforeseen addition of one or two students

[16] If more able students are more likely to leave lower performing schools, then one may conclude strategic instruction exists when it does not. However, the first stage probit model includes ITBS scores as independent variables. The unreported coefficients on ITBS scores are jointly significant and negative indicating that higher scoring ITBS students are less likely to be excluded from the analysis.

Please cite this article in press as: Krieg, J.M. Which students are left behind? The racial impacts of the No Child Left Behind Act. *Economics of Education Review* (2011), doi:10.1016/j.econedurev.2011.02.004

may make that school accountable under NCLB. Since these schools were automatically classified as AYP schools, this research may again understate the actual impacts of strategic instruction.[17]

At the time of its passage, one of the stated goals of NCLB was to eliminate the achievement gap between students of different races. NCLB may accomplish this in an unintended manner by reducing the performance of children in successful racial groups. However, shifting of resources from students of successful racial groups to less successful ones is not necessarily an inefficient use of resources. If prior to NCLB schools over-allocated resources towards a particular racial group (perhaps as a result of successful parental lobbying), then strategic instruction may result in a more efficient allocation of resources. Further, while this research presents evidence that the relative positions of racial groups has been impacted by NCLB, it also documents higher WASL test scores after the enactment of NCLB. Thus, NCLB may have changed relative racial performance while simultaneously increasing overall performance.

The implications of the current structure of NCLB can be significant for the futures of schools and society. Schools which focus their attention on poorly performing racial groups run the risk of reducing performance in their high performing racial groups. Over a period of time, it is possible that these schools will find that they have inadequately prepared students in these groups for success on high stakes test at later grades. In short, these schools may trade AYP today for their district's middle and high schools future failure when the fourth graders advance. This becomes especially important as the required pass rate increases and the performance of all students, even those in passing racial groups, becomes more critical in determining a building's AYP. For society, it is not clear that transferring resources from one racial group to another is a costless endeavor. If, for instance, schools change curricula to better engage students in an at-risk racial group, members of that group may improve but perhaps by less than members of the other groups deteriorate. The gains made by one group may or may not compensate for the losses suffered by others.

Simple alterations to NCLB could prevent this type of strategic instruction and maintain its focus on reducing racial disparities. For instance, rather than measuring the percent of each racial group that passes the WASL, NCLB could measure year-to-year average test gains by subgroup and then require each subgroup to demonstrate some appropriate amount of gains. Such a system would limit the incentive to focus on poorly performing racial groups at the expense of highly performing ones.

References

Brown, A. B., & Clift, J. W. (December 2010). The unequal effect of Adequate Yearly Progress: Evidence from school visits. *American Education Research Journal*, 744–773.

Campbell, D. T. (1976). *Assessing the impact of planned social change*. Hanover, NH: The Public Affairs Center, Dartmouth College.

Chakrabarti, R. (2007). *Vouchers, public school response and the role of incentives: Evidence from Florida*. Federal Reserve Bank of New York. Staff Reports, #306.

Cullen, J. B., & Reback, R. (2006). *Tinkering toward accolades: School gaming under a performance accountability system*. NBER Working Paper #12286.

Deere, D., & Strayer, W. (2001). *Putting schools to the test: School accountability, incentives, and behavior*. Private Enterprise Research Center, Texas A&M University, Working Paper #113.

Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, *90*, 837–851.

Figlio, D. N., & Getzler, L. S. (2002). *Accountability, ability and disability: Gaming the system*. NBER Working Paper #9307.

Jacob, B. A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in Chicago Public Schools. *Journal of Public Economics*, *89*, 761–796.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. *Educational Finance and Policy*, *3*, 250–281.

Neal, D., & Schanzenbach, D. W. (2007). *Left behind by design: Proficiency counts and test-based accountability*. NBER Working Paper #13293, 2007.

Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, *92*, 1394–1415.

Rouse, C., Hannaway, J., Goldhaber, D., & Figlio, D. (December, 2007). *Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure*. NBER Working Paper 13681.

Springer, M. G. (2008). The influence of NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*(5), 556–563.

Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J., Roby, A., & McCombs, J. S. (2008). *Pain and gain: Implementing the No Child Left Behind in California, Georgia, and Pennsylvania, 2004–2006*. The RAND Corporation.

Winters, M. A., Trivitt, J., & Greene, J. (February 2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review*, *29*(1), 138–146.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT Press.

[17] Incidentally, the requirement of 30 students per demographic group introduces another potential basis of strategic instruction. It is possible that district administrators shift school boundaries or busing routes to purposefully keep individual schools from reaching the 30 student level in weaker demographic groups.

Please cite this article in press as: Krieg, J.M. Which students are left behind? The racial impacts of the No Child Left Behind Act. *Economics of Education Review* (2011), doi:10.1016/j.econedurev.2011.02.004