The Intradistrict Effect of Educational Accountability on School Finance Under No Child Left Behind

Thomas Goldring*

November 29, 2016

Abstract

This paper examines the impact of educational accountability under the No Child Left Behind Act on the intradistrict distribution of school finance in Texas public schools. Rising performance standards over time alter the risk that a school fails to meet target proficiency rates on students' standardized exams. I use the resulting variation in the risk of failing to make Adequate Yearly Progress (AYP) to identify the causal effect of accountability pressure on changes in per pupil school expenditure and staff inputs. Compared to schools likely to make AYP, the total operating expenditure for schools likely to fail AYP are $140 per pupil less after one year (2008 dollars). The effect rises to $215 for Title I schools. Moreover, the change in total expenditure for schools on the margin to fail AYP are approximately half the magnitude of those likely to fail, suggesting that policymakers impose somewhat less stringent financial penalties for schools that are closer to meeting AYP. The effects are substantial relative to an average annual change in per pupil expenditure of $217, and robust to restricting the sample to larger districts and for estimates over two years. Changes in instructional expenditure primarily account for spending differences across schools. I also find evidence of distributional changes in staffing, including a relative increase in the total number of teachers at schools below or on the AYP margin, which is caused in part by the hiring of additional special education teachers. (JEL H52, H75, I21, I22, I28, J45)

*I thank Karen Clay, Matthew Eisenberg, Dennis Epple, Seth Richards-Shubik, and participants at the Applied Microeconomics Seminar at Carnegie Mellon University for helpful comments. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090023 to Carnegie Mellon University. The opinions expressed are my own and do not represent the views of the Institute or the U.S. Department of Education.

1 Introduction

Fifteen years after the No Child Left Behind Act (NCLB) mandated the nationwide implementation of performance-based school accountability in the U.S., the underlying mechanisms that drove student outcomes remain opaque. A growing literature has documented meaningful, if targeted, positive effects of school accountability under NCLB on student achievement (summarized by, e.g., Figlio and Ladd 2015), as well as unintended responses ranging from targeting students on the proficiency margin (Neal and Schanzenbach 2010) to outright cheating (Jacob and Levitt 2003). A number of mediating factors have been proposed to explain the causal chain between school accountability and student achievement, including shifts in time from whole-class instruction to specialist teachers in high stakes subjects (Reback et al. 2014), increased per pupil spending, higher teacher compensation, and a larger proportion of elementary school teachers with advanced degrees (Dee et al. 2013).

In this study, I examine whether school accountability under NCLB affected the distribution of per pupil spending between schools in Texas. Using nationwide district-level data, Dee et al. (2013) find evidence that NCLB increased spending by about $570 per pupil (2009 dollars). Yet accountability rewards and sanctions were focused at the school-level, thereby incentivizing school-level responses to accountability pressure. I investigate whether accountability pressure affected the school-level allocation of discretionary spending within districts. The question is important because most recent studies that use rigorous causal methods find that the level and distribution of school finance matters. Sustained increases in funding, caused by state school finance reforms, improved student performance on standardized tests, increased the probability of graduating from high school, enrolling in college, and earning a degree, and boosted hourly wages (e.g., Hyman 2013, Jackson et al. 2014, Jackson et al. 2016, Lafortune et al. 2016). If a premise of educational accountability is to improve student outcomes, then it is important to understand whether distributional changes in school finance are aligned with that goal.

My empirical strategy identifies the effect of accountability pressure by classifying schools according to their likelihood of making Adequate Yearly Progress (AYP).[1] Since actual AYP status is likely to be endogenous to changes in school expenditure, I use

[1] AYP is a binary outcome introduced under NCLB to measure school and district accountability. Section 3.2 describes how AYP was determined at the school-subgroup level in Texas.

performance on schoolwide standardized tests from prior to the determination of both actual AYP status and the level of school expenditure, in order to predict the probability that schools will make AYP. I then classify schools into three groups: (i) at low risk of failing AYP—that is, above the AYP margin, (ii) likely to fail AYP, or below the AYP margin, and (iii) on the AYP margin.[2] To estimate the effect of accountability pressure on financial outcomes, I exploit plausibly exogenous changes over time in the risk of failing AYP caused by rising proficiency standards and natural year-to-year variation in test scores. This variation allows me to compare differences in outcomes for cohorts of students that faced different degrees of accountability pressure within the same district, similar to a difference-in-differences design.

Compared to schools at low risk of failing AYP—that is, above the AYP margin—I find that operating expenditures for schools below the AYP margin are approximately $140 per pupil less after one year (2008 dollars). The effect rises to $215 for Title I schools. Moreover, total expenditure in schools on the margin to fail AYP are approximately half the magnitude of those likely to fail, suggesting that policymakers impose somewhat less stringent financial penalties for schools that are closer to meeting AYP. The effects are substantial relative to an average annual change in per pupil expenditure of $217, and robust when restricting the sample to larger districts and for estimates over two years. Differences in instructional expenditure primarily account for spending differences across schools, with only modest impacts on school leadership expenditure. Separated by program type, expenditure rises for special education students and falls for bilingual/ESL education in schools at risk of failing or likely to fail AYP, relative to schools likely to pass AYP.[3] I also find evidence of staffing changes, including a relative increase in the total number of teachers at schools below the AYP margin or on the AYP margin, which is caused in part by the hiring of additional special education teachers.

Few prior studies have examined the impact of test-based accountability on school finance, and all but one focus on accountability systems prior to NCLB. The consensus is that accountability pressure leads to changes in school spending, although the direction and type of change varies. Craig et al. (2015) find schools with higher ratings in Texas between 1994 and 2002 received more funds than lower rated schools, with the additional

[2] Reback et al. (2014) and Deming et al. (2016) employ similar empirical strategies to gauge the impact of accountability pressure, respectively, nationwide under NCLB and in Texas during the late 1990s.

[3] ESL is English as a second language.

funds targeted toward administration and training, counseling, and extra-curricular activities. Furthermore, after adopting a new statewide test, a drop in accountability rating led to a 1.5 percent increase in instructional budgets, although the increase disappeared within 3 years (Craig et al. 2013). Using four case studies, Hannaway et al. (2002) find that two states that implemented standards-based reform and accountability systems in the 1990s increased expenditures and allocated a substantial portion of the additional funds to instructional expenses (Kentucky and Texas), while the remaining two states (Maryland and North Carolina) did not. State accountability prior to NCLB also saw novice teacher salaries rise relative to veteran teachers within the same district (Bifulco 2010). The introduction of accountability systems did not affect financial equity between districts, although the introduction of a consequential accountability system did lead to improvements in resource adequacy (Rubenstein et al. 2008).[4]

The only study to my knowledge that links school accountability under NCLB with financial outcomes is by Dee et al. (2013), who use a comparative interrupted time series (CITS) design to compare outcome trends in states with accountability systems in place prior to the passage of NCLB against states with no prior accountability system. NCLB accountability increased district expenditures by an average of $570 per student (2009 dollars), and both direct instruction and pupil-support services increased.

This paper provides the first study of the impact of NCLB pressure on intradistrict school finance allocation. I investigate the causal relationship between accountability pressure and an array of financial and staff outcomes to shed light on the financial mechanisms by which districts respond to test-based accountability. This paper extends a small collection of studies which examine the impact of accountability pressure on schools that were at risk of AYP failure (Reback et al. 2014, Deming et al. 2016). A contrasting approach has examined the impact of actually receiving a failing designation using regression discontinuity (RD) designs (e.g., Chiang 2009, Rockoff and Turner 2010, Hemelt 2011, Cooley Fruehwirth and Traczynski 2012, Rouse et al. 2013, Chakrabarti 2014). My analyses show that an RD design is inappropriate for Texas data under NCLB because of the presence of discontinuities in the density of the forcing variable.[5] Schools that failed

[4] Resource adequacy is concerned with the extent to which funding is sufficient to provide students the opportunity to achieve specified learning outcomes; equity is concerned with fairness in the distribution of funding (Rubenstein et al. 2008, p. 3).

[5] Craig et al. (2015) also report discontinuities in the density of the forcing variable over the same time period. Therefore they use data from prior to NCLB to analyze the effect of marginal changes in

at least one component of the AYP rating often appealed and were designated as passing AYP. If the outcome of the appeal process is at least somewhat uncertain, then examining accountability pressure prior to the rating being determined avoids the pitfalls associated with an RD design.

In the next section I develop a simple model to provide insight into the intradistrict budget allocation decision facing district administrators. Section 3 outlines the context of school finance and school accountability under NCLB in Texas. Section 4 describes the construction of the dataset and Section 5 presents the empirical strategy used to identify accountability pressure. Section 6 reports the estimated effects of accountability pressure on school expenditure and an array of financial and staff outcomes. Section 7 concludes.

2 Analytical Framework

It is unclear *a priori* how educational accountability will affect the distribution of school finance. On the one hand, the adoption of an accountability system signals that student performance on standardized tests is given a relatively high weight in the evaluation of school outcomes. High performance in an accountability system may lead directly to rewards and recognition, and indirectly to higher budgets through the capitalization of test scores into housing prices (Figlio and Lucas 2004), implying that high-performing schools may be rewarded with higher budgetary increases than low-performing schools. On the other hand, NCLB accountability was based on student proficiency rates on standardized tests, incentivizing teachers and administrators to focus on students closest to the passing margin. If low-performing schools have a greater number of students close to the passing threshold than high-performing schools, then increases in school expenditure may have the greatest marginal effect in low-performing schools, leading to relatively higher budget increases at low-performing schools at the expense of high-performing schools. A third possibility is that school accountability plays no role in the distribution of discretionary funding, although this outcome is unlikely given that school ratings were publicized widely and reported in the local press, particularly during the heightened focus on school accountability in the early years under NCLB (Chakrabarti 2014).

Given the uncertainty in how educational accountability could affect the distribution of school funds, I develop a simple model to gain insight into the budgetary alloca-

accountability ratings on budgeted school expenditure.

tion decision facing district administrators.[6] The principal feature of the model is that performance-based accountability may impact the distribution of school finance since district policymakers prioritize performance on statewide exams. Following Bifulco (2010), I begin by defining the objective of district adminstrators to maximize the expected number of votes in an upcoming election. Suppose district policymakers must allocate a given amount of money, Θ, to increasing school operating budgets. They must choose the amount by which to increase the total operating expenditure of low-performing schools, m_L, and the amount by which to raise the budget of high-performing schools, m_H.[7] The objective function facing district policymakers may be written as

$$V = V\{P(m_L, m_H), X(m_L, m_H)\}, \tag{1}$$

where V is the expected votes received by district policymakers, P measures the increase in student achievement as measured by proficiency rates on statewide standardized tests, and X captures other factors that influence vote totals. These other factors, X, may include the support of the local teachers' union, improvements in aspects of school quality other than test performance, adherence to community norms, or any other factors that affect the expected number of total votes. P and X are each functions of the budget increase provided for, respectively, low-performing schools, m_L, and high-performing schools, m_H. P enters the objective function separately from X because under performance-based accountability, increases in school budgets may have different effects on student achievement than the other factors that influence vote totals.

Let L be the number of low-performing schools and let H be the number of high-performing schools. District policymakers maximize expected votes, V, subject to the budget constraint for increasing school operating budgets, $m_L L + m_H H = \Theta$. Assuming that $V(\cdot)$ is additively separable, the relevant first order condition is:

$$\frac{\partial V/\partial P(\partial P/\partial m_L) + \partial V/\partial X(\partial X/\partial m_L)}{\partial V/\partial P(\partial P/\partial m_H) + \partial V/\partial X(\partial X/\partial m_H)} = \frac{L}{H}. \tag{2}$$

[6] The model is very similar to that used by Bifulco (2010), who examines how performance-based accountability policies influence the distribution of teacher salary increases.

[7] The empirical analysis assigns schools to one of three groups – unlikely to fail accountability (high-performing), likely to fail (low-performing), and on the margin to fail. For ease of exposition, the model uses two groups, but without loss of generality since one of the groups may be defined as a composite increase in total operating expenditure.

If V is convex in m_H and m_L, then Figure 1 depicts the choice facing the district policymakers, where V is the "isovote" curve, Θ is the budget constraint with slope $-L/H$, and m_L^* and m_H^* are increases in the operating budgets for, respectively, low-performing schools and high-performing schools that maximize the expected number of votes. I assume that all of the partial derivatives in Equation 2 are non-negative. In particular, performance-based accountability is likely to increase the marginal effect of student achievement on expected votes, $\partial V/\partial P$, if school accountability assumes a more prominent place in voters' calculations.

The distribution of the rise in schools' total operating expenditure is determined by the relative magnitude of the partial effects of budget rises on student achievement. An increase in the marginal effect of raising the budget of high-performing schools on student achievement, $\partial P/\partial m_L$ makes the isovote curve steeper (V to V_1 in Figure 1), which leads to a larger increase in the expenditure of high-performing schools, m_H^1, and a lower increase in the budget of low-performing schools, m_L^1. Conversely, an increase in the marginal effect of raising the expenditure of low-performing schools on student achievement makes the isovote curve flatter and leads to a larger increase in the budget of high-performing schools, m_H^*, relative to low-performing schools, m_L^*. The empirical analyses in this study amount to a statistical test of which of the two marginal effects on student achievement assumes a greater role under school accountability.

3 Policy Background

The public education system in Texas is one of the largest in the U.S. By 2008, there were 1,253 school districts (the most in the nation) with 8,195 schools in operation, of which 374 were charter schools.[8] A small number of districts—294—comprised a single school; the median number of schools within a district was 3 while the largest district operated 292 schools (Houston ISD). The system enrolled 4.7 million students (behind only California) and employed 633,000 school employees (the most nationwide). In what follows, I outline salient features of the school finance and performance-based accountability systems in Texas during the NCLB era.

[8] The analysis sample in this paper ends with test score data from the 2007-08 academic year.

3.1 School Finance

School districts in Texas are funded through a combination of local (property) tax, state funds, and federal funds. In the 2007-08 academic year, elementary-secondary revenue totaling $46.1 million was divided between $21.7 million (47 percent) raised from local taxes, $19.9 million (43 percent) from state funds, and $4.5 million (10 percent) from federal funding sources (U.S. Census Bureau 2010).

The majority of a district's funding entitlement is determined by school funding formulae, which are primarily based on weighted average daily attendance. Districts may also levy an additional local tax to boost revenues, although voter approval is required for high rates. Additional funds are provided for facilities expenditures (Texas Taxpayers and Research Association 2012). After a series of landmark decisions by the Texas Supreme Court,[9] in which the court found that the state's method of funding public schools discriminated against students in poorer districts, the legislature attempted to promote a more equitable interdistrict distribution of revenue by capping the property tax rate used for maintenance and operations (M&O). Furthermore, revenues above a statewide rate were either 'recaptured' by the state and distributed to poorer districts, or transferred from wealthier to poorer districts under an agreement between districts.

School funds are distributed within districts by a combination of formula-driven state aid, which is targeted towards students who are economically disadvantaged or have limited English proficiency (LEP), and discretionary spending at the district level (Ajwad 2006). The observed level of per pupil spending represents a combination of formula-driven state aid and the discretionary allocation of school funds. Intradistrict discretionary spending in Texas is primarily, although not unilaterally, progressively distributed with respect to poverty rates (Ajwad 2006, Baker 2012).

Figure 2 displays mean school expenditures per pupil by academic year. The mean total operating expenditure declined from approximately $6,300 per pupil in 2002-03 to $6085 in 2005-06, followed by an increase to $6,410 in 2007-08.[10] Despite the spending

[9] The Texas Supreme Court case is Edgewood ISD v. Kirby, case numbers 777 S.W.2d 391 (TX 1989) and 777 S.W.2d 491 (TX 1991).

[10] Figure 2 displays expenditure data for schools in the analysis sample, which excludes vocational schools and schools that focus on educating special education students, charter schools, schools in districts with less than 3 schools, and schools with fewer than 10 students. All school finance data in the analysis sample are in 2008 dollars. The expenditure data in 2001-02 is budgeted spending; by contrast, spending from 2002-03 to 2007-08 is actual school expenditure. Thus the sharp increase in the mean per pupil expenditure between 2001-02 and 2002-03 is likely an artifact of the sample construction and may not be as pronounced

rise after 2006, Texas ranked 42nd in per pupil spending in 2007-08 (U.S. Census Bureau 2010). Two of the components of total spending—instructional expenditure per pupil and leadership expenditure per pupil—are also shown in Figure 2; they both also fell and then rose between 2002-03 and 2007-08, such that the per pupil expenditure levels were similar in 2008 as in 2003. The pertinent question in this paper is whether the discretionary intradistrict allocation of the change in school funds shown in Figure 2 was responsive to differential accountability pressure between schools.

3.2 School Accountability Under NCLB

At the time of NCLB's enactment in 2001, Texas had had a consequential accountability system in place since 1994. The state accountability system, signed into law by Governor George W. Bush, was in many respects a model for NCLB, and Texas continued to use its state accountability system to rate schools based on student performance alongside the AYP ratings mandated by NCLB.[11] AYP statuses are publicly available (online and on school report cards) and AYP failure was associated with negative publicity as well as potential stigma and shame (Chakrabarti 2014). Moreover, schools that received Title I funding and failed AYP repeatedly were subject to an increasing set of sanctions.[12]

Under NCLB, districts and schools were required to meet criteria on three measures to make AYP: reading/language arts,[13] mathematics, and one other measure. All students and each subgroup (black, Hispanic, white, students with disabilities, economically dis-

a rise if comparable time series were used. Unfortunately, actual school expenditure data for 2001-02 are unavailable.

[11] The first state accountability system in Texas was used from 1994 to 2002, and its centerpiece features included annual student tests (grades 3-8 and 10) on state exams, publicly releasing average test results, and assigning districts and schools into performance categories—low-performing (school), acceptable (school), academically unacceptable (district), academically acceptable (district), recognized (district/school), and exemplary (district/school)—based on test results in reading, writing, and mathematics, as well as attendance and high school dropout. Average test scores were also calculated separately by four subgroups—white, African-American, Hispanic, and economically disadvantaged—but only if the subgroup comprised at least 10 percent of the school's number of test takers. A second state accountability system was used from 2004 to 2011. It reduced the number of rating labels to four and made them consistent across districts and schools, evaluated test results from grades 3-11, and evaluated the completion rate for high schools.

[12] A school in its second consecutive year failing AYP must offer school choice within the district to its students. In the third consecutive year, schools must additionally provide supplemental educational services to low-income families. In the fourth consecutive year, the district must either replace school staff, implement a new curriculum, decrease management authority at the school, extend the school day or school year, appoint an outside expert, or reorganize the school internally.

[13] Henceforth I refer to reading/language arts as 'reading.'

advantaged, limited English proficient) that met minimum size criteria had to meet either the performance standard or performance gains criteria, and the participation standard (90 percent). The performance standards for academic years 2002-03 through 2007-08 are displayed in Figure 3. The standards increased every two years to meet NCLB's requirement that schools should be 100 percent proficient by 2014. If any student subgroup (or the all students group) failed to meet the performance standard for reading and mathematics, the group could still make AYP through performance gains criteria (also known as 'safe harbor'): both (i) a 10 percent decrease in the percent not passing the standards on tests from the prior year, and (ii) any improvement on the other measure. The minimum size criteria for inclusion in the AYP performance calculation were 50 or more students and the subgroup had to comprise at least 10 percent of all test takers in the subject, or 200 or more students even if the subgroup represented less than 10 percent of all test takers in the subject.

The other measure was either the graduation rate for high schools and any campuses offering grade 12, or the attendance rate for elementary schools, middle/junior high schools, and any school not offering grade 12. Districts and schools had to meet the graduation rate performance standard (70 percent) or the attendance rate performance standard (90 percent) for the whole school, or show any improvement from the prior year. The other measure was calculated for individual student subgroups solely for the purpose of applying safe harbor. Overall, a school could be evaluated on as few as 2 or as many as 29 measures to determine AYP status. Each school received a binary overall AYP status; failure on any applicable subgroup resulted in the school receiving the label "Missed AYP."

4 Data

To link school accountability to financial outcomes, I construct a new longitudinal dataset that combines test proficiency statistics, binary AYP outcomes, and school characteristics data at the school-subgroup level, and school-level data on financial expenditure and staffing outcomes.[14] I supplement incomplete AYP outcomes data for student subgroups from the Texas Education Agency (TEA) with individual school AYP reports and publicly-

[14] All expenditure data are in 2008 dollars, the final year of the analysis sample.

available AYP data used by Reback et al. (2014).[15] I also use school characteristics from the Common Core of Data (CCD) compiled by the National Center for Education Statistics (NCES). Further details on the data collected are presented in Appendix A.

I examine two groups of outcomes in response to accountability pressure: (i) changes in per pupil expenditures, and (ii) changes in staffing and classroom characteristics, which clarify how spending changes affect school resources. The data come from the Academic Excellence Indicator System (AEIS), a comprehensive annual school-level database maintained by TEA. In particular, I obtain annual data on total campus expenditure per pupil and two components of total spending: instructional and leadership expenditure. I also examine expenditure outcomes by program type, including bilingual/ESL, gifted and talented, and special education. The staffing outcomes include teacher counts for total teachers, regular education FTE, and special education FTE. For classroom characteristics, I examine students per teacher, average teacher salary, and the number of educational aides.

I restrict the analysis sample by removing (i) vocational schools and schools that focus primarily on educating special education students, according to school type classifications in the Common Core of Data, (ii) charter schools, (iii) schools in districts with less than 3 schools, because my research question focuses on intradistrict changes in discretionary spending, and (iv) small schools that enrolled less than 10 students.[16] I test the robustness of the results with respect to the district size restriction by reanalyzing all models after restricting the analysis sample to large districts with at least 6 schools. All exclusion restrictions are applied on a year-by-year basis, so that a school excluded in one year may be present in the analysis sample in another time period.

Since the performance standards for AYP increased biennially, I examine the effect of accountability pressure over two years in which the performance standards remained constant. The first year under NCLB was 2002-03 and the performance standards remained constant across 2002-03 and 2003-04. Thus I construct measures of accountability pressure

[15] I use AYP outcomes data from Reback et al. (2014) for 2002-03 and 2003-04, the first two full academic years under NCLB. I scrape AYP outcomes data at the school-subgroup level from individual school AYP reports for the next three years, from 2004-05 to 2006-07; Appendix A provides further detail about this process. Starting in 2007-08, annual school-subgroup AYP outcomes data is available in test proficiency data supplied by TEA.

[16] I use the restriction of removing small schools which enrolled less than 10 students because AYP classifications in Texas were made on a case-by-case basis for schools with less than 10 students. Small schools could also be evaluated using a pairing relationship with another campus.

using data from the prior academic year, 2001-02, and examine outcomes in 2002-03 and 2003-04. Henceforth I refer to the two-year interval as a 'cohort'. I measure accountability pressure for two further cohorts – first, for 2004-05 and 2005-06, and, second, for 2006-07 and 2007-08. As Figure 3 shows, the AYP performance standards remained constant by cohort. The final year in my analysis sample is 2007-08 for several reasons: (i) the AYP performance standards rose annually rather than biennially after 2008, (ii) TEA adopted a growth measure for accountability—the Texas Projection Measure—in 2008-09 that markedly reduced the percentage of campuses that were labeled as missing AYP, and (iii) the Obama administration announced that states could apply for waivers in September 2011. Thus the academic years from 2002-03 to 2007-08 represent the most stable period of time for the Texan accountability system under NCLB in its original form.

The number of student subgroups that counted towards AYP varied widely between schools. Table 1 displays the percentage of schools in which each student subgroup was numerically significant. Economically disadvantaged students were most likely to count for accountability purposes in more than 80 percent of schools. The largest racial and ethnic subgroups were Hispanic and white students; the Hispanic group was numerically significant in 70 percent of schools by 2008, up from 64 percent in 2004, while the white subgroup trended down from 65 percent to 60 percent of schools. Other subgroups (African American, special education, and LEP students) were numerically significant in 21 percent to 33 percent of schools.

Tables 2–4 provide summary statistics for, respectively, control variables, per pupil expenditure, and staffing and classroom characteristics used in the analysis sample. As Table 2 reports, campuses are about 14 percent African-American, 44 percent Hispanic, and 38 percent white. 46 percent of students are eligible for free or reduced-price lunch.[17] 68 percent of schools received Title I funds and therefore were subject to escalating federally-mandated sanctions if the school persistently failed AYP.

The mean total operating expenditure per pupil was $6,281 (2008 dollars), with a one-year change of $217 and a two-year change of $262 (Table 3). The largest component of total expenditure per pupil by school function was instructional expenditure, which averaged $4,452 (71 percent). School leadership expenditure accounted for $436 (7 percent). By school program, special education spending accounted for $771 per pupil, bilingual

[17] Students qualify for a free or reduced-price meal if their family income is less than 185 percent of the Federal poverty line.

instruction for $257, and gifted and talented $93. The total number of teachers was 66, with 45 regular education teachers and 7 special education teachers. The mean total number of teachers increased over one year by 0.7 teachers and by 0.9 teachers over two years. There were 15 students per teacher within a classroom and the mean fell a little over time. Teachers earned on average $43,400, which increased by $1,570 over one year and $2,475 over two years. On average, there were about 10 educational aides at each school.

5 Econometric Specification

In an ideal world, the causal effect of NCLB accountability on school finance could be measured by randomly assigning AYP status across schools and estimating the resulting change in school expenditure. AYP status is in fact determined (for the most part) by student performance on standardized exams, which is also likely to affect the level of, and year-to-year change in, school expenditure. Since a school's actual AYP status is thus likely to be endogenous to changes in school finance, I instead measure the effect of accountability pressure using a two-step process. First, I predict the likelihood that each student subgroup will meet the performance standards under NCLB based on prior student achievement on statewide tests. I then classify schools into one of three groups based on the probability of making AYP across all subgroups that count towards AYP: on the margin for making AYP, above the margin, or below the margin.

In the second step, I use the AYP margin designation to examine differences in school finance and staffing outcomes within the same district between schools above, on, or below the AYP margin. In practice, the omitted category is schools above the AYP margin, so all effects of accountability pressure are relative to schools likely to pass AYP. This method is similar to that used by Reback et al. (2014) to examine the effect of accountability pressure based on variation in state performance standards. The key difference is that I compare schools within districts that must meet rising proficiency standards over time, while Reback et al. (2014) compare similar schools across states in which proficiency standards vary widely.[18]

[18] In addition to identifying variation from rising proficiency standards over time, there is also yearly variation in prior test scores, which may be due to differences in the demographic composition of the student body. Since changes in the demographic composition of the student body may be endogenous if parental enrollment decisions respond to performance-based accountability, I include the demographic controls shown in Table 2 in every model specification.

In the remainder of this section, I describe in detail the procedure used to measure accountability pressure. First, for each subject s, I estimate regressions of the following form:

$$\mathbf{AYP}_{jksc} = \begin{cases} 1 & \text{if } \alpha_q + \mathbf{X}_{jks,c-1}\beta_1 + \beta_2 \mathrm{N}_{jksc} + \mathbf{X}\mathrm{N}_{jks}\beta_3 \\ & \quad + \mathbf{W}_{j,c-1}\beta_4 + \mathbf{M}_{jksc}\beta_5 + \zeta_{jks} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\mathbf{AYP}_{jksc}$ denotes whether subgroup k at school j met its AYP proficiency rate targets in subject s for cohort c. $\mathbf{X}_{jks,c-1}$ is a vector containing a cubic polynomial for subgroup test performance on the Texas Assessment of Knowledge and Skills (TAKS) standardized test during the academic year prior to the two-year cohort, and N_{jksc} is a student subgroup size variable for cohort c, which is entered into the model as the inverse of the square root of the number of accountable test-taking students in subject s in subgroup k in school j. $\mathbf{X}\mathrm{N}_{jks}$ represents interaction terms between prior test scores and the student subgroup size variable, $\mathbf{W}_{j,c-1}$ is a vector of control variables for school-level demographics during the academic year prior to the two-year cohort, $\mathbf{M}_{jksc}$ is a vector of indicators for whether the subgroup was numerically significant in only one of the two cohort years, and ζ_{jks} is a normally distributed latent disturbance term. Since the error variance in student pass rates decreases mechanically as the number of tested students increases, I include subgroup size variables (N_{jksc}) and interactions with test score variables ($\mathbf{X}\mathrm{N}_{jks}$).

To measure accountability pressure, I classify schools by the degree of pressure each school faced under NCLB. Since schools failed AYP based on the performance of the lowest performing subgroup, schools with at least one subgroup unlikely to meet the performance standards were highly likely to fail AYP. Similarly, schools in which all subgroups had a high likelihood of meeting the performance standards were highly likely to make AYP. Based on this logic, Reback et al. (2014, p. 218) established the following definitions to measure accountability pressure under NCLB:

1. A school is classified as *above the AYP margin* if all numerically significant subgroups have a high chance of making AYP in both math and reading.

2. A school is classified as *below the AYP margin* if it has at least one numerically significant subgroup with a low chance of making AYP in either math or reading.

3. A school is classified as *on the AYP margin for a particular subject* if at least one

numerically significant subgroup in the school has a moderate chance of making AYP in that subject, and no numerically significant subgroup in the school has a low chance of making AYP in either subject.

4. A school is classified as *on the AYP margin* if it is on the AYP margin for math or reading.

In other words, schools above the AYP margin are likely to make AYP, schools below the AYP margin have a high probability of failing AYP, and the potential outcome for schools on the AYP margin is uncertain. Following Reback et al. (2014), I define a 'high' chance of a subgroup making AYP as above 75 percent, a 'moderate' chance as between 25 percent and 75 percent, and a 'low' chance as less than 25 percent. The results are not sensitive to varying the high probability cutoff between 65 and 85 percent, or varying the low probability cutoff between 15 and 35 percent.

A majority of schools are classified as above the AYP margin in all three cohorts (see Table 2), ranging from almost 70 percent in 2004 to 84 percent in 2006. The percentage of schools rated as highly likely to fail AYP—that is, below the AYP margin—fell from 6.5 percent in 2004 to 2.5 percent in 2006, then rose to almost 8 percent of schools in 2008. The most likely explanation for the fall in 2006 is that certain schools were exempted from some accountability requirements during the 2005-06 academic year in the aftermath of Hurricane Katrina; many schools in Texas experienced an influx of students who had fled affected areas in Louisiana. The dip in the percentage of schools rated as below the AYP margin in 2006 is offset by a rise in the percentage classified as above the margin; the proportion on the AYP margin is largely unaffected.

Figure 4 displays the change in schools' positions on the AYP margin between 2004 and 2008. Of schools below the AYP margin in 2004, just under 60 percent remained below the AYP margin in 2008, while 30 percent were on the AYP margin and 10 percent were above the AYP margin in 2008. For schools on the AYP margin in 2004, over 10 percent were subsequently below the AYP margin—that is, likely to fail AYP—in 2008. About 25 percent remained at risk of failing in 2008, while just over 60 percent were likely to pass AYP. Conversely, 90 percent of schools with high probability of passing AYP in 2004 were also likely to pass in 2008. The variation in AYP status used to identify the effect of accountability pressure comes largely from schools that were initially likely to fail AYP or at risk of failing. Figures 5 and 6 display the change in AYP status from, respectively, the cohort in 2004 to 2006, and the cohort in 2006 to 2008. The data show that schools were likely to improve their position on the AYP margin between 2004 and 2006, due in part to the accountability exemptions put in place after Hurricaine Katrina. Conversely, schools on the AYP in 2006 were almost equally likely to be rated as below the AYP margin or on the margin in 2008. Overall, the data indicate the presence of plausibly exogenous variation in accountability pressure over time, which I use to compare schools within districts by cohort.[19]

I estimate the impact of accountability pressure on various outcomes with the following regression specification:

$$Y_{jc} = \delta_1 \hat{M}_{jc} + \delta_2 \hat{B}_{jc} + \mathbf{W}_{jc}\delta_3 + \alpha_d + \gamma_c + \epsilon_{jc}, \tag{4}$$

where Y_{jc} is an outcome of interest, $\hat{M}_{jc}$ and $\hat{B}_{jc}$ are indicators for schools, respectively, on the AYP margin and below the AYP margin, $\mathbf{W}_{jc}$ is a vector of control variables for school-level demographics, α_d are district fixed effects, γ_c are cohort fixed effects, and ϵ_{jc} is a normally distributed latent disturbance term. The variables of interest are δ_1 and δ_2, which represent the average impact of accountability pressure under NCLB for schools on or below the AYP margin. The coefficients have a causal interpretation if the variation in accountability pressure across cohorts is exogenous. I test this assumption by conducting a variety of robustness checks. First, I compare estimates of accountability pressure on the subsequent change in school expenditure under NCLB with estimates for a falsification sample using spending data from the 1990s, prior to the introduction of NCLB. If predicted accountability pressure is exogenous, I expect to find no effect on predetermined outcomes. Second, I estimate Equation 4 again using only data from larger districts (with at least 6 schools). Third, I impute missing subgroup data using the proficiency rate for all students and recalculate the accountability pressure predictions as well as all outcome models.

[19] I examine whether the AYP margin classifications are plausibly exogenous by comparing my results against a falsification sample. The falsification sample contains the same expenditure outcomes but for the years prior to NCLB, which coincides with the statewide accountability system Texas had in place prior to NCLB. If the schools that are low-performing in the years prior to NCLB's adoption are the same as the schools below the AYP margin or at risk of failing under NCLB, then the falsification sample may be biased in favor of statistically significant effects from accountability pressure. However, since the majority of the estimates based on the falsification sample are not significant at conventional levels, it provides evidence in support of the assumption that the measure of accountability pressure based on predicted probabilities captures exogenous variation in pressure.

6 Results

Panel A of Tables 5–8 displays results for the analysis sample, which includes all districts with at least 3 schools. I start by examining the change in expenditure per pupil by school function. The first column in Table 5 displays estimated effects of accountability pressure on total operating expenditure over one year. Compared to schools above the AYP margin, the change in total operating expenditure for schools on the AYP margin or below the AYP margin are, respectively, \$83.05 per pupil and \$139.56 per pupil less over one year. Both differences are statistically significant ($p < 0.01$ in both cases) and large considering that the average change in total operating expenses per pupil over one year was \$217 (Table 3). The change in total operating expenses over two years are very similar in magnitude to the change over one year, which suggests that district officials responded promptly to accountability pressure (Table 5, column 2). Point estimates for the falsification sample are not statistically significant, the one-year estimates have opposing signs, and the two-year effects are in the opposite direction, lending credence to the claim that the empirical strategy identifies exogenous accountability pressure.

Title I schools that persistently failed AYP were subject to an escalating series of sanctions. Therefore I restrict the main analysis sample to Title I schools to examine whether sanction-driven accountability pressure differed from accountability pressure with no sanctions. The effects of accountability pressure are stronger in Title I schools than for schools without sanctions attached; the change in total operating expenditure for schools on the AYP margin or below the AYP margin are, respectively, \$128.79 per pupil and \$215.21 per pupil less over one year (Table 5, column 1, panel B). The effects are both statistically significant at the 0.01 level. I also restrict the main analysis sample to districts with at least 6 schools to check whether accountability pressure varies by district size. Compared to the main analysis sample, there is little difference in the magnitude of the change in total operating expenditure for schools on the AYP margin. The magnitude is somewhat less for schools below the AYP margin, at \$108.18 per pupil less over one year (Table 5, column 1, panel C). The effects over two years are similar in magnitude to one-year impacts. Figure 7 displays estimated effects on the one-year change in total operating expenditure as the minimum number of schools within each district increases. The estimates are robust to the restriction of only using larger districts and statistically significant regardless of the minimum number of schools.

Two spending components of total operating expenses are instructional expenditure (the largest component) and school leadership expenditure. The third and fourth columns in Table 5 display estimated effects of accountability pressure for instructional expenditure, and the fifth and sixth columns display impacts for school leadership spending. The change in instructional expenditure for schools on the AYP margin and below the AYP margin tell a similar story to total operating expenditure, albeit with a smaller magnitude (which is unsurprising since instructional expenditure is a subset of total operating expenditure). The effects persist for Title I schools, in larger districts, and across one or two years. Similarly, for the change in school leadership expenditure, schools below the AYP margin had a larger drop compared to schools on the AYP margin, and both effects were less than schools above the AYP margin. The effects for schools on the AYP margin were only marginally statistically significant for the main analysis sample and for schools in larger districts, although they were significant at the 5 percent level under the restriction of Title I schools. Moreover, the estimates for the falsification are not significant (except for one estimate over two years for instructional expenditure which differs in sign from the main estimates) and do not show a consistent pattern of effects in any direction or by magnitude. Overall, the point estimates for school expenditure are consistent in direction, magnitude, across time, and when restricted to Title I schools or larger districts.

Table 6 displays estimated effects of accountability pressure on the change in expenditure by program type. Estimates in the first column show that, compared to schools above the AYP margin, the changes in spending on bilingual/ESL education for schools on the AYP margin or below the AYP margin are, respectively, \$9.67 per pupil and \$15.72 per pupil less over one year. Both estimates increase in magnitude to, respectively, \$20.69 per pupil and \$37.49 per pupil for the change in spending over two years. The results are qualitatively similar when the main analysis sample is restricted to Title I schools or small districts. The falsification sample estimates are significant at the 10 percent level for schools below the AYP margin, although the magnitude is much less than for the main analysis sample. It is possible that similar patterns of effects occurred under Texas' prior accountability system, which are causing the significant results here, or that the results are in line with one or two estimates for the falsification sample being significant by chance. The pattern and direction of the falsification estimates, though, suggests that the estimates using the main analysis sample should be interpreted with a degree of caution.

Columns 3 and 4 of Table 6 show estimated effects on the change in spending for gifted and talented education; the direction of point estimates varies between models and estimates are close to 0, suggesting that accountability pressure had no effect on gifted and talented education. The fifth and sixth columns report estimates for the change in spending for special education students. Point estimates are uniformly positive and the difference in the change in spending over two years is statistically significant and large considering that the average change in special education spending was $26 (Table 3). The falsification estimates are also positive but smaller in magnitude and none are statistically significant at conventional levels.

Given accountability-driven distributional changes in spending, I explore shifts in school staffing and classroom characteristics. The estimates suggest that the number of teachers rose more in schools below the AYP margin (Table 7, columns 1 and 2). Compared to schools above the AYP margin, the change in the average number of teachers for schools on the AYP margin or below the AYP margin rose by, respectively, 0.3 ($p = 0.11$) and 0.9 ($p = 0.02$) teachers. Interestingly, the effects for the subset of staff teaching regular education classes do not show corresponding increases; the point estimates are almost uniformly negative for the change over one year and mostly positive over two years, but none of the estimates are statistically significant (Table 7, columns 3 and 4). Conversely, there is some evidence that special education teachers rose more in schools on the AYP margin and below the AYP margin—by 0.1 ($p = 0.05$) and 0.3 ($p = 0.05$) teachers, respectively—suggesting that the relative increase in the number of teachers was concentrated among those teaching special education rather than in regular education classrooms (Table 7, columns 5 and 6).

Table 8 documents estimates for three additional classroom indicators. For schools on the AYP margin or below the AYP margin, the student-teacher ratio was relatively higher one year after accountability pressure and increased further over two years (Table 8, columns 1 and 2). The student-teacher ratio increased by 0.022 after one year and by 0.088 ($p = 0.05$) after two years for schools on the AYP margin, although the one-year estimate is not statistically significant. The average teacher salary increased by less for schools on the AYP margin ($44.10 less after one year and $61.47 less after two years) or below the AYP margin ($106.42 less after one year and $101.30 less after two years) compared to those above the AYP margin, although the estimates are not statistically significant (Table 8, columns 3 and 4). The picture is less clear for the change in the number of educational aides in the classroom, although there is some evidence that the number of aides rose for schools on the AYP margin and decreased for schools below the AYP margin, one year after accountability pressure (Table 8, columns 5 and 6). Interestingly, this is the only outcome for which the sign of the effect varies between schools on the AYP margin and schools below the AYP margin, although most estimates are not statistically significant. Thus there is some evidence that school administrators may have responded differently to the type of accountability pressure when deciding upon staffing needs for educational aides.

Lastly, I test the robustness of the results to the imputation of missing prior subgroup proficiency rates, in order to avoid dropping observations in the logit models. I impute missing subgroup data using the proficiency rate for all students and recalculate predicted accountability pressure and all outcome models. The results are presented in Tables B1–B4, and the estimates are uniformly similar in magnitude and statistical significance to the estimates from the main analysis sample presented herein.

7 Conclusion

The No Child Left Behind Act introduced a nationwide performance-based accountability system in which public schools were designated as making or failing AYP. Recent evidence suggests that the AYP rating affected student achievement through a variety of mediating factors (Murnane and Papay 2010, Dee et al. 2013, Reback et al. 2014), and this paper contributes to that literature by showing that accountability pressure under NCLB affected the distribution of changes in school finance. Total operating expenditures for schools likely to fail AYP and at risk of failing are, respectively, approximately $83 and $140 per pupil less after one year (2008 dollars) compared to schools likely to pass AYP, and the effects persist two years after accountability pressure is measured. The effects are larger for Title I schools which face increasing sanctions for persistent failure. Both instructional expenditure and school leadership expenditure also fall for low-performing schools and schools on the AYP margin, relative to high-performing schools.

The distributional changes in school expenditure have hetergeneous effects on spending by program type. Expenditure rises for special education students and falls for bilingual/ESL education in schools at risk of failing or likely to fail AYP, relative to schools likely to pass. I also find evidence of staffing changes for the same group of schools, including an increase in the total number of teachers, which is driven in part by the hiring of more special education teachers, and educational aides. Point estimates for the change in teacher salaries are negative, although the estimates are not statistically significant at conventional levels.

Given the analytical framework developed in Section 2, my findings suggest that the marginal effect of raising the budget of high-performing schools on student achievement is greater than the marginal effect of raising the budget of low-performing schools. This finding is important because it suggests that policymakers are using AYP as a measure of school quality to either reward high-performing schools for good performance or sanction low-performing schools. This echoes the results of Craig et al. (2015), who find that schools just over the rating boundary received more financial resources than schools just under the boundary. Yet in contrast to their focus on schools close to the rating cutoff through the use of an RD design, I examine the entire distribution of schools according to their probability of making AYP. I find that all schools on the AYP margin are on average financially worse off, at least compared to high-performing schools.

Although the outcomes in this study are from the first six years under NCLB, the findings are relevant as states design new accountability systems to comply with the Every Student Succeeds Act (ESSA). While AYP is eliminated under ESSA, states must design accountability systems based on results on standardized tests in reading, math, and science, English language learner proficiency, one other academic measure, and at least one nonacademic measure. The lowest-performing five percent of all schools must be identified as "in need of improvement." If district policymakers would like to give the lowest-performing schools that same resources afforded to high-performing schools, then discretionary funds should be carefully allocated to achieve an equitable distribution of school expenditure.

References

Ajwad, Mohamed Ihsan. 2006. "Is Intrajurisdictional Resource Allocation Equitable?: An Analysis of Campus-Level Spending Data for Texas Elementary Schools." *The Quarterly Review of Economics and Finance*, 46(4): 552–564.

Baker, Bruce D. 2012. "Rearranging Deck Chairs in Dallas: Contextual Constraints and Within-District Resource Allocation in Urban Texas School Districts." *Journal of Education Finance*, 37(3): 287–315.

Bifulco, Robert. 2010. "The Influence of Performance-Based Accountability on the Distribution of Teacher Salary Increases." *Education Finance and Policy*, 5(2): 177–199.

Chakrabarti, Rajashri. 2014. "Incentives and Responses Under No Child Left Behind: Credible Threats and the Role of Competition." *Journal of Public Economics*, 110: 124–146.

Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics*, 93(9): 1045–1057.

Cooley Fruehwirth, Jane and Jeffrey Traczynski. 2012. "Spare the Rod? The Dynamic Effects of No Child Left behind on Failing Schools." Unpublished.

Craig, Steven G., Scott A. Imberman, and Adam Perdue. 2013. "Does It Pay to Get an A? School Resource Allocations in Response to Accountability Ratings." *Journal of urban Economics*, 73(1): 30–42.

Craig, Steven G., Scott A. Imberman, and Adam Perdue. 2015. "Do Administrators Respond to Their Accountability Ratings? The Response of School Budgets to Accountability Grades." *Economics of Education Review*, 49: 55–68.

Dee, Thomas S., Brian Jacob, and Nathaniel L. Schwartz. 2013. "The Effects of NCLB on School Resources and Practices." *Educational Evaluation and Policy Analysis*, 35(2): 252–279.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment and Earnings." *The Review of Economics and Statistics*, 98(5).

Figlio, David N. and Helen F. Ladd. 2015. "School Accountability and Student Achievement." In: Helen F. Ladd and Margaret E. Goertz (editors), "Handbook of Research in Education Finance and Policy," pages 194–210. Routledge, New York, NY.

Figlio, David N. and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 94(3): 591–604.

Hannaway, Jane, Shannon McKay, and Yasser Nakib. 2002. "Reform and Resource Allocation: National Trends and State Policies." In: William J. Fowler, Jr. (editor), "Developments in School Finance, 1999–2000," pages 57–76. U.S. Department of Education, National Center for Education Statistics, Washington, DC.

Hemelt, Steven W. 2011. "Performance Effects of Failure to Make Adequate Yearly Progress (AYP): Evidence from a Regression Discontinuity Framework." *Economics of Education Review*, 30(4): 702–723.

Hyman, Joshua. 2013. "Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment." Unpublished.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2014. "The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement, and Adult Outcomes." National Bureau of Economic Research.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *The Quarterly Journal of Economics*, 131(1): 157–218.

Jacob, Brian A. and Steven D Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics*, 118(3): 843–877.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2016. "School Finance Reform and the Distribution of Student Achievement." National Bureau of Economic Research.

Murnane, Richard J. and John P. Papay. 2010. "Teachers' Views on No Child Left Behind: Support for the Principles, Concerns about the Practices." *The Journal of Economic Perspectives*, 24(3): 151–166.

Neal, Derek and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *The Review of Economics and Statistics*, 92(2): 263–283.

Reback, Randall, Jonah Rockoff, and Heather L. Schwartz. 2014. "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools Under No Child Left Behind." *American Economic Journal: Economic Policy*, 6(3): 207–241.

Rockoff, Jonah and Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy*, 2(4): 119–147.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy*, 5(2): 251–281.

Rubenstein, Ross, Sonali Ballal, Leanna Stiefel, and Amy Ellen Schwartz. 2008. "Equity and Accountability: The Impact of State Accountability Systems on School Finance." *Public Budgeting & Finance*, 28(3): 1–22.

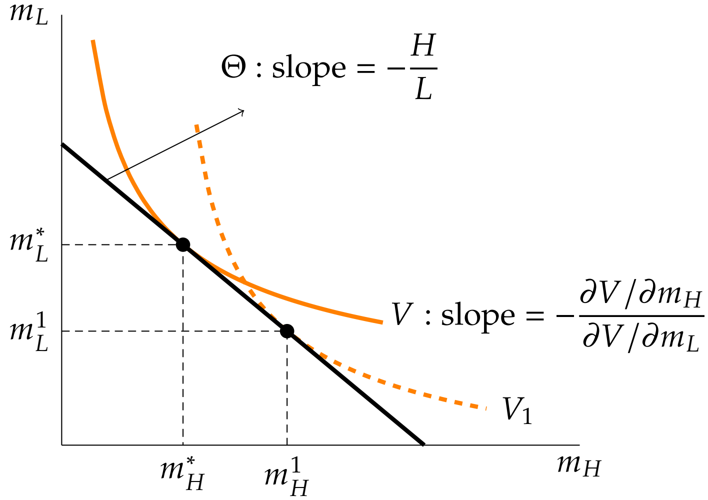
Texas Education Agency. 2010. "Texas Consolidated State Application Accountability Workbook." Technical report. `https://rptsvr1.tea.texas.gov/ayp/txworkbook10.pdf`.

Texas Taxpayers and Research Association. 2012. "An Introduction to School Finance in Texas." Technical report. `http://www.ttara.org/files/document/file-4f1732f763446.pdf`.

U.S. Census Bureau. 2010. "Public Education Finances 2008." Technical report. `http://www2.census.gov/govs/school/08f33pub.pdf`.

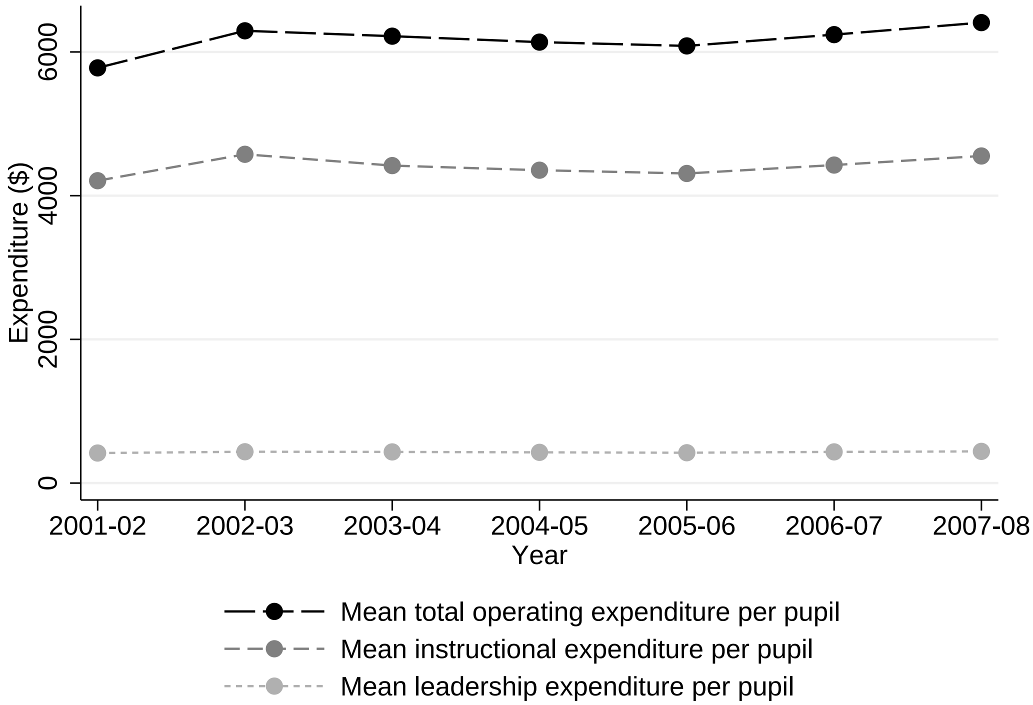
Figures & Tables

Figure 1: Distribution of budgeted increases in schools' operating expenditure



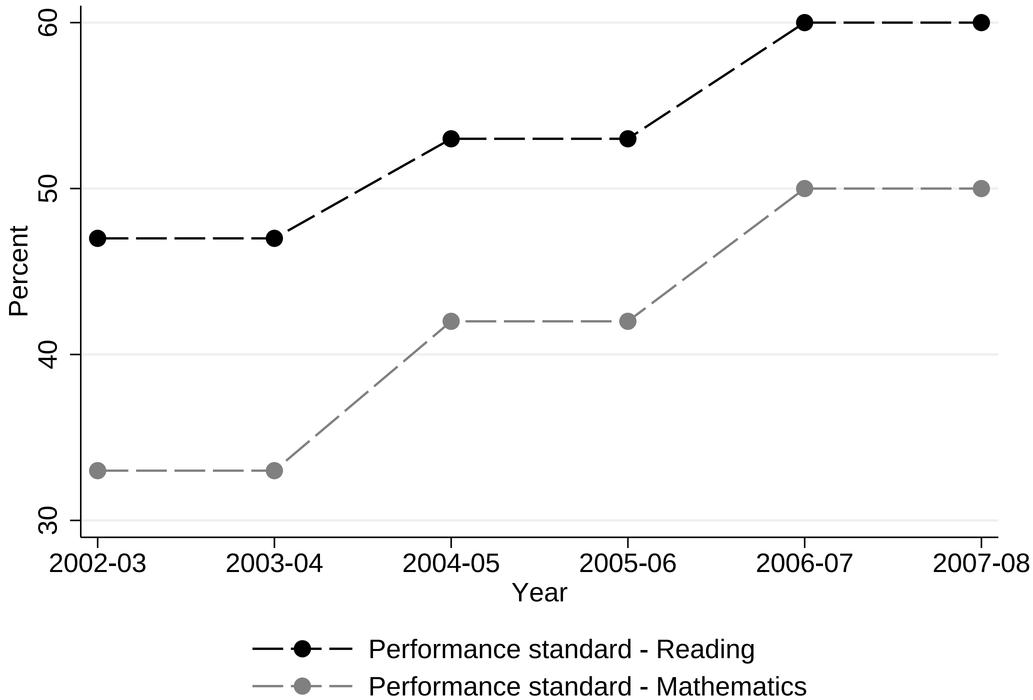
Notes: The variables are as defined in the text.

Figure 2: Mean school expenditures per pupil in Texas, by academic year



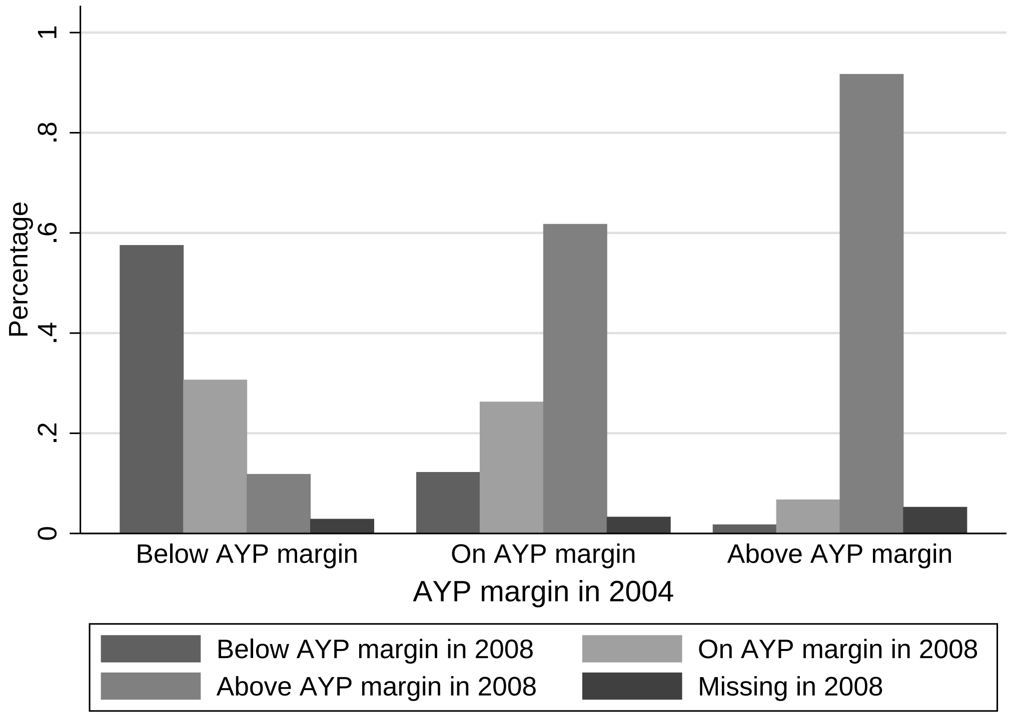
Notes: The data source is the Academic Excellence Indicator System (AEIS) at TEA. The data shown excludes vocational schools and schools that focus primarily on educating special education students, according to school type classifications in the Common Core of Data. Mean expenditures are calculated as the weighted average of school-level per pupil expenditures, where the weights are the total number of students by academic year.

Figure 3: AYP performance standards under NCLB in Texas, by academic year



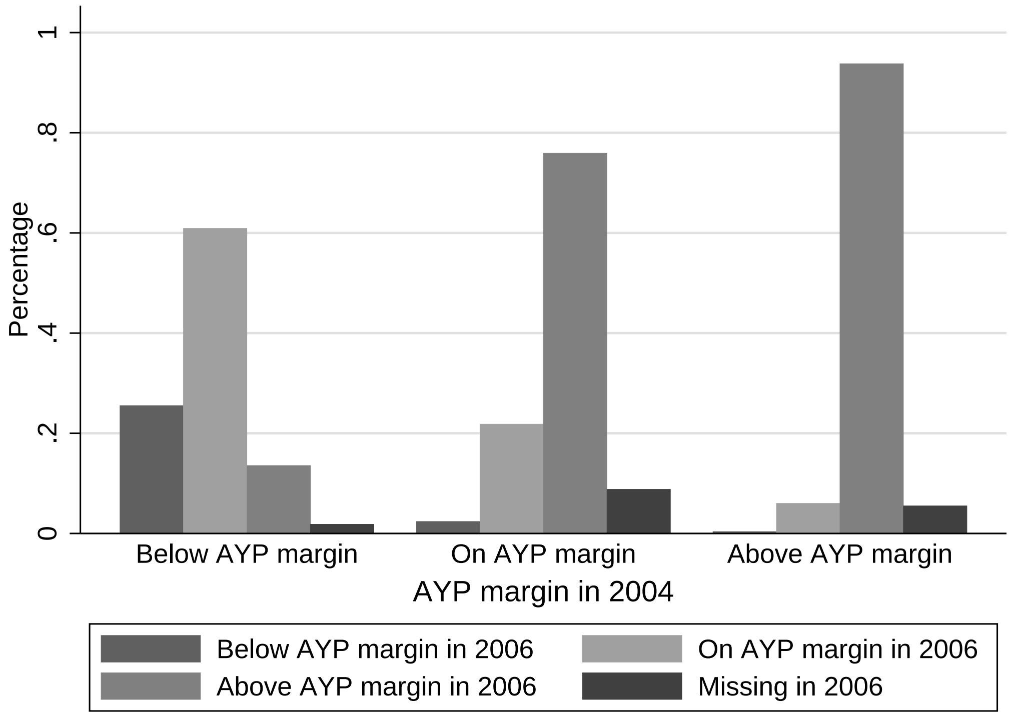
Notes: The data source is the Texas Education Agency (2010, p. 10).

Figure 4: Change in AYP margin between 2004 and 2008



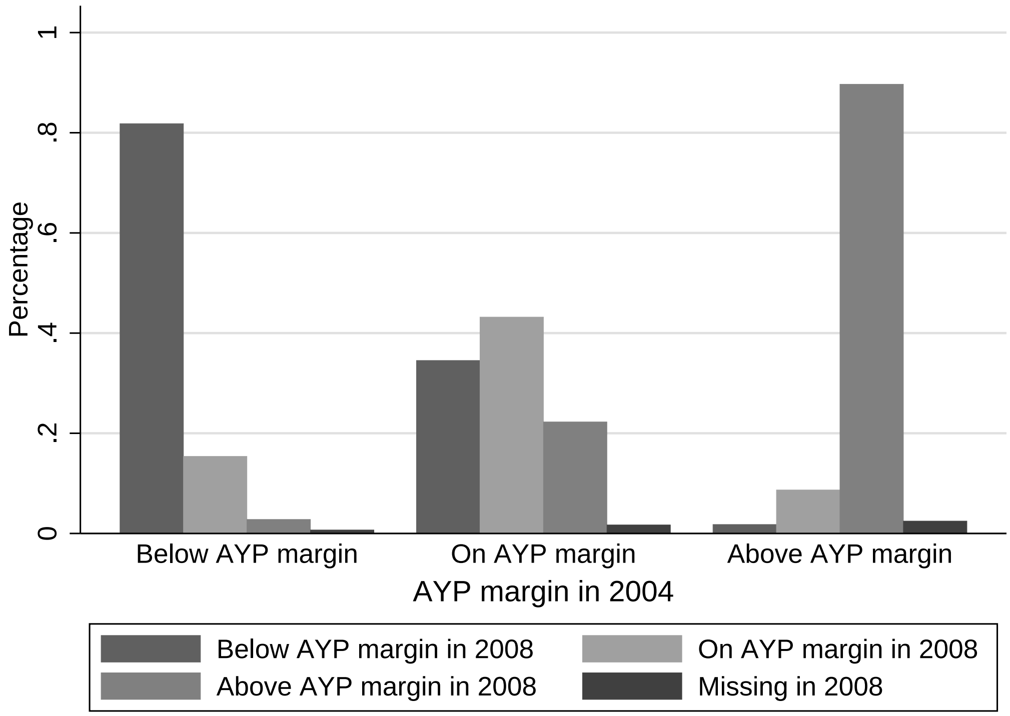
Notes: See the text for details on the data and procedure used to construct the pressure prediction.

Figure 5: Change in AYP margin between 2004 and 2006



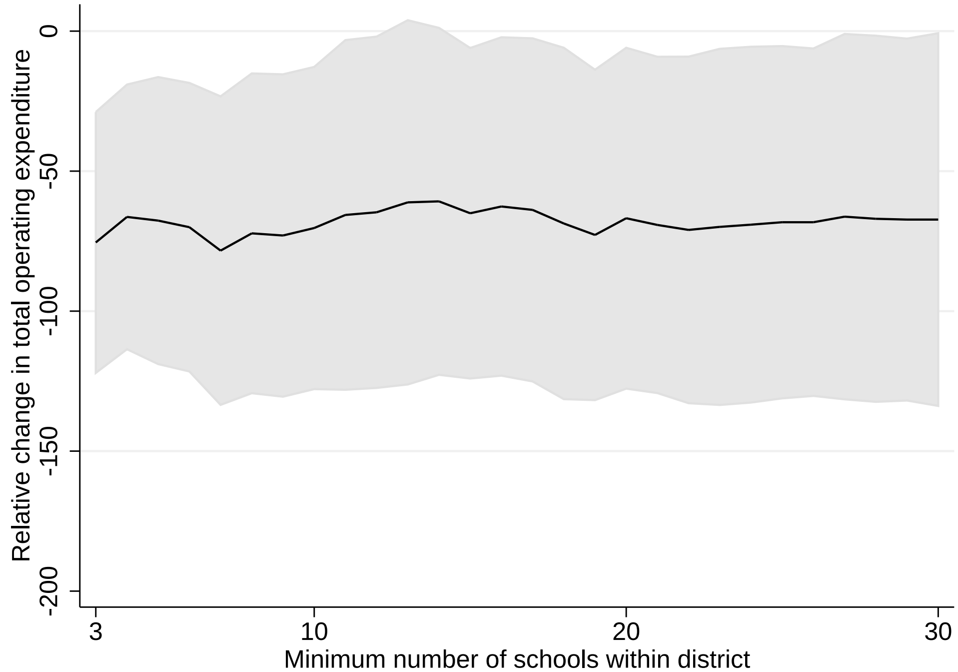
Notes: See the text for details on the data and procedure used to construct the pressure prediction.

Figure 6: Change in AYP margin between 2006 and 2008

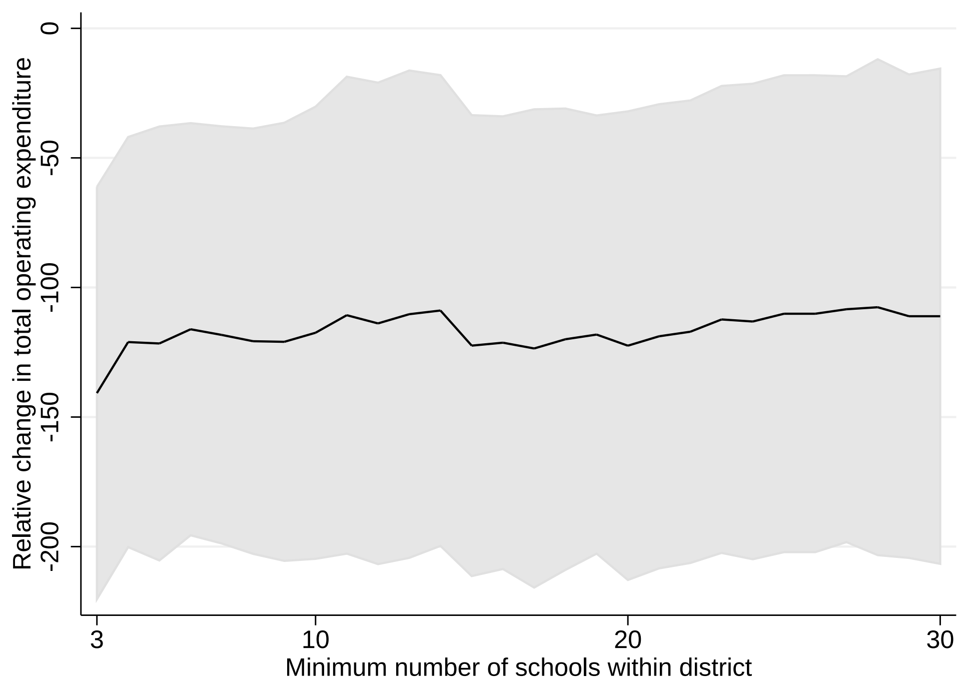


Notes: See the text for details on the data and procedure used to construct the pressure prediction.

Figure 7: Effect of accountability pressure on the change in total operating expenditure



(a) On AYP margin



(b) Below AYP margin

Table 1: Numerically significant subgroups by year

	Numerically significant subgroup		
	2003-2004	2005-2006	2007-2008
Overall school population	100.0%	100%	100%
African American	24.9%	26.2%	27.0%
Hispanic	64.2%	67.5%	70.6%
White	64.8%	62.8%	60.0%
Economically disadvantaged	80.7%	83.5%	84.5%
Special Education	26.5%	27.4%	21.1%
Limited English proficient	29.7%	31.6%	32.9%

Notes: The table shows the percentage of schools in the analysis sample in which the student subgroup was numerically significant and counted as part of the AYP determination. The subgroup definitions are those used by the Texas Education Agency (TEA) to classify students and students may appear in more than one subgroup. The data source is the Academic Excellent Indicator System maintained by TEA.

Table 2: Descriptive statistics for school control variables and measure of accountability pressure

	Mean	Std. Dev.	*N*
Control variables			
Black	14.2%		17,782
Hispanic	44.0%		17,782
White	38.4%		17,782
Free or reduced price lunch	46.2%		17,782
Title I	68.0%		17,782
Total students	673.6	482.9	17,782
Prior reading proficiency rate	88.2%	8.2%	17,640
Prior math proficiency rate	82.2%	14.3%	17,719
Accountability pressure			
Below AYP margin	5.7%		17,571
2003-04	6.5%		5,863
2005-06	2.5%		5,680
2007-08	7.8%		6,028
On AYP margin	16.9%		17,575
2003-04	24.4%		5,863
2005-06	13.4%		5,682
2007-08	13.0%		6,030
Above AYP margin	77.6%		17,782
2003-04	69.1%		5,864
2005-06	84.3%		5,791
2007-08	79.5%		6,127

Notes: The table shows mean statistics calculated using school-level data in the analysis sample. Control variables (with the exception of Title I and total students) are weighted by the total number of students at the school.

Table 3: Descriptive statistics for school expenditure

	Mean	Std. Dev.	*N*
Per pupil expenditure, by school function			
Total operating	\$6,281	\$1,128	17,643
Change over 1 year	\$217	\$724	17,635
Change over 2 years	\$262	\$832	17,635
Instruction	\$4,452	\$730	17,643
Change over 1 year	\$155	\$583	17,635
Change over 2 years	\$152	\$647	17,635
School leadership	\$436	\$126	17,643
Change over 1 year	\$9	\$69	17,635
Change over 2 years	\$12	\$84	17,635
Per pupil expenditure, by school program			
Bilingual	\$257	\$501	17,643
Change over 1 year	\$10	\$139	17,625
Change over 2 years	\$28	\$202	17,635
Gifted & talented	\$93	\$182	17,643
Change over 1 year	\$0	\$70	17,628
Change over 2 years	\$5	\$92	17,635
Special education	\$771	\$390	17,643
Change over 1 year	\$26	\$166	17,635
Change over 2 years	\$99	\$255	17,635

Notes: The table shows mean statistics calculated using school-level data in the analysis sample. Expenditures are weighted by the total number of students at the school.

Table 4: Descriptive statistics for school staffing outcomes

	Mean	Std. Dev.	*N*
School staffing			
Total teacher count	66.1	46.0	17,760
Change over 1 year	0.7	6.1	17,756
Change over 2 years	0.9	9.4	17,760
Regular education FTE	44.9	30.5	17,760
Change over 1 year	0.6	6.6	17,756
Change over 2 years	0.4	8.9	17,760
Special education FTE	6.7	6.0	17,760
Change over 1 year	0.1	2.0	17,756
Change over 2 years	0.2	2.6	17,760
Classroom characteristics			
Students per teacher	15.2	2.0	17,760
Change over 1 year	−0.1	1.1	17,755
Change over 2 years	−0.2	1.4	17,760
Average teacher salary	$43,396	$4,309	17,760
Change over 1 year	$1,570	$1,650	17,755
Change over 2 years	$2,475	$2,128	17,760
Educational aides	10.2	6.0	17,760
Change over 1 year	0.0	2.7	17,756
Change over 2 years	0.0	3.5	17,760

Notes: The table shows mean statistics calculated using school-level data in the analysis sample. Expenditures are weighted by the total number of students at the school.

Table 5: Effects of AYP pressure on expenditure per pupil, by function

Dependent variable:	Total operating		Instruction		School leadership	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	−83.05***	−87.23***	−60.96***	−67.12***	−4.56*	−5.90*
	(22.63)	(26.97)	(18.05)	(21.18)	(2.60)	(3.09)
Below AYP margin	−139.56***	−153.64***	−115.14***	−124.48***	−13.76***	−17.25***
	(40.06)	(54.08)	(30.33)	(40.10)	(4.60)	(5.94)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	−128.79***	−133.28***	−90.85***	−94.05***	−7.93**	−10.32**
	(29.35)	(36.91)	(22.29)	(27.72)	(3.56)	(4.27)
Below AYP margin	−215.21***	−238.73***	−182.53***	−200.23***	−13.22**	−18.27**
	(51.18)	(71.69)	(39.08)	(54.14)	(6.54)	(8.37)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	−71.96***	−76.31**	−51.50***	−59.10***	−4.96*	−5.51*
	(24.94)	(30.96)	(18.97)	(22.27)	(2.70)	(3.23)
Below AYP margin	−108.18***	−143.06***	−92.61***	−111.30***	−9.09*	−14.73**
	(38.69)	(54.36)	(30.08)	(40.98)	(4.65)	(6.14)
Panel D. Falsification sample						
On AYP margin	−13.38	55.21*	−5.30	42.76*	−1.51	2.10
	(19.00)	(30.74)	(15.55)	(22.52)	(2.35)	(3.49)
Below AYP margin	29.10	75.32	−22.47	47.64	0.93	1.89
	(32.31)	(55.63)	(24.57)	(38.75)	(3.83)	(6.40)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A are 17,407; for Panel B are 11,872; and for Panel C are 13,340. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in expenditure over one or two academic years.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table 6: Effects of AYP pressure on expenditure per pupil, by program

Dependent variable:	Bilingual/ESL		Gifted & Talented		Special Education	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	−9.67*	−20.69***	−0.76	−0.67	12.32**	30.11***
	(5.08)	(6.47)	(2.85)	(3.78)	(5.32)	(8.53)
Below AYP margin	−15.72**	−37.49***	−0.11	−4.95	13.38	56.08***
	(6.22)	(9.07)	(4.68)	(5.62)	(9.31)	(14.67)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	−14.13*	−27.57***	−0.25	0.06	8.78	22.17*
	(7.96)	(9.88)	(3.28)	(4.69)	(7.24)	(11.31)
Below AYP margin	−25.08**	−50.76***	−1.81	−8.42	10.99	60.51***
	(9.83)	(13.91)	(4.67)	(5.79)	(12.60)	(20.57)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	−10.41*	−22.74***	−0.82	−1.05	13.34**	31.32***
	(6.30)	(8.21)	(3.09)	(4.18)	(5.85)	(9.55)
Below AYP margin	−18.62**	−43.73***	0.11	−5.31	17.18*	54.41***
	(7.20)	(10.16)	(4.70)	(5.93)	(9.06)	(14.92)
Panel D. Falsification sample						
On AYP margin	−1.04	−8.18	0.53	1.82	5.49	8.80
	(3.47)	(5.27)	(1.99)	(3.34)	4.17	(7.54)
Below AYP margin	−5.47*	−11.76*	0.14	−3.63	7.58	9.93
	(4.17)	(6.64)	(3.07)	(5.22)	6.34	(11.26)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range between 17,397 and 17,407; for Panel B are between 11,866 and 11,872; and for Panel C are between 13,333 and 13,340. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in expenditure over one or two academic years.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table 7: Effects of AYP pressure on school staffing

Dependent variable:	Total Teacher Count		Regular ed. FTE		Special ed. FTE	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	0.253	0.473**	−0.094	0.124	0.111*	0.073
	(0.159)	(0.233)	(0.180)	(0.237)	(0.057)	(0.076)
Below AYP margin	0.932**	1.368**	−0.283	−0.010	0.276**	0.373*
	(0.382)	(0.567)	(0.428)	(0.557)	(0.139)	(0.204)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	0.331*	0.616**	0.002	0.217	0.126**	0.017
	(0.193)	(0.244)	(0.237)	(0.300)	(0.063)	(0.078)
Below AYP margin	1.082**	1.569**	−0.236	0.014	0.284*	0.233
	(0.442)	(0.665)	(0.514)	(0.672)	(0.161)	(0.225)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	0.239	0.445	−0.176	0.015	0.132*	0.118
	(0.203)	(0.281)	(0.218)	(0.299)	(0.069)	(0.094)
Below AYP margin	0.864**	1.232**	−0.422	−0.208	0.288*	0.385*
	(0.391)	(0.623)	(0.454)	(0.592)	(0.148)	(0.211)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range between 17,526 and 17,530; for Panel B are between 11,954 and 11,957; and for Panel C are between 13,445 and 13,446. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in staffing counts over one or two academic years.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 8: Effects of AYP pressure on classroom characteristics

Dependent variable:	Students per teacher		Avg. teacher salary		Educational Aides	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	0.022	0.088*	−44.10	−61.47	0.173**	0.155
	(0.037)	(0.047)	(38.94)	(49.60)	(0.084)	(0.107)
Below AYP margin	0.055	0.157	−106.42	−101.30	−0.232	0.035
	(0.071)	(0.097)	(70.19)	(86.24)	(0.161)	(0.220)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	0.073	0.152**	−31.71	−67.52	0.152	0.085
	(0.050)	(0.064)	(51.46)	(66.41)	(0.111)	(0.140)
Below AYP margin	0.139	0.317**	−41.37	−56.94	−0.257	0.007
	(0.099)	(0.136)	(91.63)	(123.58)	(0.214)	(0.288)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	0.025	0.082	−35.24	−36.15	0.180*	0.174
	(0.042)	(0.053)	(42.55)	(53.39)	(0.098)	(0.122)
Below AYP margin	0.047	0.173*	−58.45	−58.80	−0.236	0.022
	(0.073)	(0.103)	(72.97)	(90.14)	(0.171)	(0.239)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range from 17,525 to 17,530; for Panel B are between 11,953 and 11,957; and for Panel C are between 13,444 and 13,446. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in staffing counts over one or two academic years.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

A Data Sources

Figure 3 shows that the AYP performance standards in Texas increased every two years during the first 6 years under NCLB. As described in the paper, I predict the probability of failing AYP during each two-year block in which the AYP performance standards remained fixed. I use test scores in 2002 to predict the likelihood of failing AYP in 2003 and 2004; I use test scores in 2004 to predict AYP in 2005 and 2006; and I use test scores in 2006 to predict AYP in 2007 and 2008. The predictions are calculated at the school-subgroup level and then combined to assess whether a school is below, on, or above the margin to make AYP (see the paper for more detail). In the remainer of this section, I provide additional details on the data sources and coding used to construct the dataset in this paper.

A.1 Subject-specific subgroup AYP proficiency indicator

In the probit models that predict the probability of failing AYP, the dependent variable is a subject-specific subgroup AYP proficiency indicator. Unfortunately, the subgroup-level proficiency indicator is not publicly available from TEA for the first 5 years of NCLB. (A school-level AYP proficiency indicator is available.) I use three data sources to obtain subgroup-level AYP proficiency indicators from 2003-2008. First, for 2003 and 2004, I use AYP indicators from Reback et al. (2014), which they in turn requested from TEA as part of their comprehensive project to collect AYP proficiency indicators and related performance metrics from every state for the first two years of NCLB. Second, for 2005 to 2007, I collected a two-page campus data report, in pdf format, for every campus in Texas.[20] Each data report contains performance metrics on standardized tests (TAKS), attendance, graduation, and test participation, and, crucially, subject-specific subgroup AYP proficiency indicators. I converted each campus data report from pdf format to Microsoft Excel format (.xlsx).[21] I then cleaned up the campus data report in Excel using a Visual Basic for Applications (VBA) script, retaining only the AYP proficiency indicators. I merged together in Excel the indicators from every campus data report and finally imported the data into Stata.

[20] I wrote a batch script using the automation language AutoIt to download every report for each academic year.

[21] I used an API in R provided by PDFTables (`https://www.pdftables.com`).

A.2 Prior subgroup test performance

To predict the probability of failing AYP for the 2002-04 cohort, I use the subgroup-specific proficiency rate on the Texas Assessment of Academic Skills (TAAS) standardized test from 2002. TAAS was replaced by the Texas Assessment of Knowledge and Skills (TAKS) in 2003, so I use subgroup-specific proficiency rates on TAKS from 2004 and 2006 to predict the probability of failing AYP for, respectively, the 2004-06 and 2006-08 cohorts. The data source is the Academic Excellence Indicator System (AEIS) maintained by TEA.

The TAAS in 2002 was not broken out for students with limited English proficiency (LEP), one of the subgroups used to determine AYP status. To avoid dropping the LEP subgroup, I use the campus-wide proficiency rate as the LEP proficiency rate, and I also create a variable measuring the percent of all students that the LEP subgroup accounted for in 2002. This variable is equal to 1 if the actual subgroup proficiency rate is used. I create a second proficiency variable identical to the first but in which I fill in all missing subgroup proficiencies with the campus-wide proficiency rate (not only LEP). I also create a second variable measuring the percent of all students that the filled in subgroups accounted for. This variable is equal to 1 if the actual subgroup proficiency rate is used.

Prior proficiency rates are entered into the logistic regression models with linear, quadratic, and cubic terms. The variable capturing the percent of all students that the filled in subgroups accounted for is entered as a main effect and interacted with the three prior proficiency rate terms.

A.3 Student subgroup size

The data source is the AEIS. Student subgroup size is entered as 1/sqrt(size). It is interacted with the three prior proficiency rate terms and the three prior proficiency rate × the subgroup percent interaction terms.

A.4 Control variables

The school-level control variables include: the percent of students who are black, Hispanic, and white; the percent of students who qualify for free or reduced-price meal; an indicator for Title I eligibility status; and the total count of students. The data source for all control variables is the AEIS, with the exception of the Common Core of Data for FRL.

B Additional Tables

Table B1: Effects of AYP pressure on expenditure per pupil, by function, with imputation for missing subgroup proficiency rates

Dependent variable:	Total operating		Instruction		School leadership	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	−86.90***	−96.68***	−60.42***	−70.62***	−4.95**	−6.62**
	(22.49)	(26.60)	(17.88)	(20.94)	(2.50)	(3.01)
Below AYP margin	−141.57***	−149.99***	−115.36***	−119.59***	−15.10***	−17.96***
	(40.83)	(55.62)	(31.52)	(40.93)	(4.50)	(5.85)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	−130.41***	−142.33***	−88.88***	−98.01***	−7.88**	−10.54***
	(28.44)	(36.49)	(21.15)	(27.75)	(3.34)	(4.00)
Below AYP margin	−218.69***	−225.45***	−182.93***	−187.06***	−14.70**	−17.40**
	(52.82)	(73.38)	(40.07)	(54.87)	(6.73)	(8.11)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	−75.72***	−85.36***	−51.31***	−63.31***	−5.27**	−5.87*
	(24.31)	(30.66)	(19.13)	(22.48)	(2.64)	(3.11)
Below AYP margin	−108.46***	−136.51**	−91.91***	−105.31**	−10.30**	−15.03**
	(39.56)	(55.21)	(31.13)	(41.95)	(4.72)	(5.95)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. Missing subgroup-subject proficiency rates are imputed using the overall school-subject proficiency rate and the model includes a term for the percent that the subgroup comprised. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A are 17,407; for Panel B are 11,872; and for Panel C are 13,340. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in expenditure over one or two academic years.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B2: Effects of AYP pressure on expenditure per pupil, by program, with imputation for missing subgroup proficiency rates

Dependent variable:	Bilingual/ESL		Gifted & Talented		Special Education	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	−10.04*	−22.70***	−0.37	0.41	15.59***	32.97***
	(5.26)	(6.62)	(2.90)	(3.80)	(5.21)	(8.42)
Below AYP margin	−16.31***	−39.39***	−0.18	−5.03	13.59	51.80***
	(6.26)	(8.96)	(4.88)	(5.82)	(9.38)	(15.06)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	−14.57*	−31.13***	0.23	1.61	8.26	21.29**
	(8.15)	(10.00)	(3.44)	(4.76)	(7.18)	(10.77)
Below AYP margin	−25.62***	−53.96***	−1.82	−7.38	7.07	50.88**
	(9.49)	(13.83)	(4.88)	(5.85)	(12.94)	(21.11)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	−10.94*	−25.15***	−0.74	−0.32	17.72***	36.00***
	(6.63)	(8.41)	(3.06)	(4.18)	(5.81)	(9.47)
Below AYP margin	−19.05**	−45.54***	−0.13	−5.70	18.17*	51.30***
	(7.41)	(10.18)	(4.91)	(6.19)	(9.33)	(15.23)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from equation 4 in the paper. Missing subgroup-subject proficiency rates are imputed using the overall school-subject proficiency rate and the model includes a term for the percent that the subgroup comprised. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range between 17,397 and 17,407; for Panel B are between 11,866 and 11,872; and for Panel C are between 13,333 and 13,340. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in expenditure over one or two academic years.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table B3: Effects of AYP pressure on school staffing, with imputation for missing subgroup proficiency rates

Dependent variable:	Total Teacher Count		Regular ed. FTE		Special ed. FTE	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	0.283*	0.507**	−0.028	0.194	0.115**	0.081
	(0.157)	(0.229)	(0.179)	(0.238)	(0.057)	(0.075)
Below AYP margin	0.865**	1.385**	−0.381	−0.069	0.263*	0.380*
	(0.399)	(0.597)	(0.437)	(0.577)	(0.150)	(0.208)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	0.346*	0.646***	0.011	0.244	0.133**	0.025
	(0.195)	(0.245)	(0.235)	(0.301)	(0.063)	(0.076)
Below AYP margin	1.053**	1.735**	−0.136	0.199	0.228	0.188
	(0.476)	(0.696)	(0.553)	(0.740)	(0.171)	(0.233)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	0.280	0.492*	−0.089	0.112	0.133*	0.119
	(0.201)	(0.279)	(0.215)	(0.292)	(0.069)	(0.094)
Below AYP margin	0.793*	1.246*	−0.523	−0.270	0.272*	0.388*
	(0.412)	(0.654)	(0.456)	(0.615)	(0.157)	(0.217)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. Missing subgroup-subject proficiency rates are imputed using the overall school-subject proficiency rate and the model includes a term for the percent that the subgroup comprised. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range between 17,526 and 17,530; for Panel B are between 11,954 and 11,957; and for Panel C are between 13,445 and 13,446. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in staffing counts over one or two academic years.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B4: Effects of AYP pressure on classroom characteristics, with imputation for missing subgroup proficiency rates

Dependent variable:	Students per teacher		Avg. teacher salary		Educational Aides	
Change over:	1 year	2 years	1 year	2 years	1 year	2 years
Panel A. All schools in districts with at least 3 schools						
On AYP margin	0.021	0.098**	−51.82	−71.02	0.153*	0.129
	(0.036)	(0.046)	(38.91)	(49.21)	(0.083)	(0.106)
Below AYP margin	0.063	0.143	−80.12	−58.41	−0.222	0.060
	(0.072)	(0.100)	(71.15)	(87.53)	(0.168)	(0.223)
Panel B. Title I schools in districts with at least 3 schools						
On AYP margin	0.066	0.164*	−42.91	−80.61	0.128	0.077
	(0.049)	(0.064)	(50.45)	(62.29)	(0.109)	(0.138)
Below AYP margin	0.152	0.286*	−17.68	−2.42	−0.190	0.110
	(0.101)	(0.140)	(93.24)	(124.56)	(0.235)	(0.296)
Panel C. All schools in districts with at least 6 schools						
On AYP margin	0.021	0.089*	−37.85	−46.07	0.158*	0.131
	(0.041)	(0.053)	(42.28)	(51.69)	(0.096)	(0.120)
Below AYP margin	0.054	0.156	−27.40	−12.74	−0.228	0.038
	(0.074)	(0.107)	(75.20)	(91.23)	(0.182)	(0.244)

Notes: Each column is a single regression of the indicated outcome for the panel on the set of variables from Equation 4 in the paper. Missing subgroup-subject proficiency rates are imputed using the overall school-subject proficiency rate and the model includes a term for the percent that the subgroup comprised. All models control for the variables listed in Table 2 and include district and year fixed effects. Bootstrapped standard errors, adjusted for school-level clustering and using 1,000 Monte Carlo simulations, are displayed in parantheses below each estimate. Sample sizes for Panel A range from 17,525 to 17,530; for Panel B are between 11,953 and 11,957; and for Panel C are between 13,444 and 13,446. Each coefficient gives the impact for a two-year student cohort of being on or below the margin for AYP on school expenditure outcomes. See the text for details on the construction of the pressure prediction. The outcomes are the change in staffing counts over one or two academic years.

** Significant at the 5 percent level.

* Significant at the 10 percent level.