Does test-based accountability improve more than just test scores?

Katherine McElroy [*]

Department of Economics, University of Illinois at Chicago, 601 S Morgan St, Chicago, IL 60607, United States

ARTICLE INFO

ABSTRACT

This paper estimates the long-run effects of school accountability on educational attainment by exploiting two sources of variation: staggered implementation of accountability across states and individuals' exposure to accountability. I find 12 years of exposure to school accountability leads to an increase in the likelihood of graduating high school by 2.3 percentage points but has no statistically significant effect on college attendance or the likelihood of receiving a Bachelor's degree. However, racial heterogeneity shows Hispanic students experience a significant increase in the likelihood of attending college. I rule out changes in school expenditures and teacher characteristics as potential mechanisms and present suggestive evidence that schools are classifying more students as learning disabled. Lastly, accountability is more effective in conjunction with promotion gates.

1. Introduction

School accountability, the system in which schools are responsible for student performance, has been at the forefront of education policy in America and other countries for over two decades. There is substantial evidence suggesting school accountability has successfully increased student test scores (Carnoy & Loeb, 2002; Dee & Jacob, 2011; Hanushek & Raymond, 2005). However, school accountability has been criticized for causing teachers to teach to the test, to teach to the students who are near the margin of passing, or to sort to well-performing schools. Evidence of student performance across the achievement distribution suggests students in the middle of the distribution benefit the most from school accountability, but students in the tails of the distribution perform worse on standardized tests than they would have without the policy (Lee & Orfield, 2006; Neal & Schanzenbach, 2010). As a result, it is still unclear whether school accountability is beneficial to student performance and educational attainment.

I examine the long-term effects of school accountability on students' high school graduation, college attendance, and college graduation. Although test scores are an acceptable proxy for measuring student learning, they have their shortcomings. For instance, it is not possible to distinguish between improved learning or improved test-taking skills nor do they offer any measure of students' soft skills. Therefore, studying the long-term outcomes is necessary as it offers insight into whether the test score results are a reflection of better test-taking skills or if they improve student learning and lead to higher educational attainment.

Using American Community Survey data for the years 2008 to 2016, I estimate the effect of school accountability on educational attainment. I utilize two sources of plausibly exogenous variation in exposure to accountability policies to estimate a dynamic differences-in-differences event study analysis. First, I take advantage of the staggered implementation of school accountability across states throughout the 1990s and early 2000s. Second, I use variation in exposure to accountability based on an individual's age when school accountability was adopted in their state. Individuals who were out of school (18 years or older) the year of or before school accountability was adopted in their state are considered untreated. Individuals who are of school-going age during or after the year the policy passed are considered treated and are assigned a value for the number of years they are exposed to school accountability. I find 12 years of exposure to school accountability leads to an increase in the likelihood of graduating high school by 2.3 percentage points on average but has no statistically significant effect on college attendance or college graduation.

As the No Child Left Behind (NCLB) Act emphasized the performance of minority students which may have led to more concentrated effects among Hispanic and Black students, I conduct a heterogeneity analysis by race. The estimates show Hispanic and Black students are not significantly more likely to graduate high school than white students. However, Hispanic students are 7.7 percentage points more likely to attend college if they were exposed to school accountability for the entirety of their school life. The outcomes of Black students are not statistically significantly different from those of White students. The

---

findings are consistent with previous heterogeneity analyses that estimate the effect of school accountability on student test scores (Carnoy & Loeb, 2002; Dee & Jacob, 2011).

Further, I investigate whether the presence of promotion gates, as defined by high school exit exams, influences the efficacy of school accountability. There may be complementarities between school accountability and promotion gates as both emphasize testing, or promotion gates might be a signal that a state has a stricter accountability policy. The results show a large part of the effect on the high school graduation rate is coming from the interaction between exposure to school accountability and exposure to promotion gates, suggesting that promotion gates in conjunction with school accountability lead to larger increases in high school graduation rates.

School accountability is a nebulous policy in which individual states have significant control over its implementation, even in the post-NCLB era. Therefore, it is important to explore what particular policy choices might be driving the results reported in this paper. I investigate school expenditures, teacher wage and qualifications, and the proportion of students eligible for special education and language assistance programs. I find suggestive evidence that the proportion of students eligible for special education increased after school accountability, implying schools may be reacting to accountability policies by more proactively testing students for special education.

Due to the national adoption of school accountability, the long-run effects of school accountability are challenging to study, and the current literature is sparse. There are only two published studies that look at the long-run effects of school accountability, Carnoy and Loeb (2002) and Deming et al. (2016). This is the first paper to investigate the general effect of school accountability, as opposed to comparing a weaker version of school accountability to a stronger version (Deming et al. 2016).[1] The general effect of school accountability is meaningful as states have significant autonomy in designing their school accountability policies, even after the implementation of NCLB, and knowing the average long-run effect of accountability on the generations of students affected by this policy may influence future policy decisions. Moreover, this paper is the first to look at the effect of school accountability by exposure.[2] This is also policy-relevant as it is important to know whether the effect of school accountability is a result of compounded exposure or if just minor exposure to the policy can induce significant effects.

2. About school accountability

State, district, and school accountability gained popularity through the 1990s and peaked in 2002 with the passage of the No Child Left Behind Act. The NCLB Act was signed into law in 2002 and established a uniform set of requirements for school accountability across all states. Before NCLB, each state had autonomy over its school accountability policy and, as a result, states' implementation of the policy varied in many aspects. For the sake of this paper, I define a state as having a school accountability policy if it has school-level performance reporting in the form of a public report card for each public school, referred to as report card accountability, or if it has school-level consequences for not meeting performance goals, referred to as consequential accountability.[3,4] After NCLB, all states had consequential accountability, therefore my main analysis excludes any state that had report card accountability at any time prior to NCLB in order to have consistent treatment for the whole time period.[5] Table 1 lists the states that implemented school accountability before NCLB, the type of school accountability that was implemented, and the date of implementation.

Because all states have to follow the guidelines of NCLB after 2002, I briefly discuss the NCLB Act (NCLB, 2002). First, NCLB mandates testing in math and reading at least once within grades 3 through 5, grades 6 through 8, and grades 10 through 12 (NCLB, 2002). Results of these tests must be reported for specific student subgroups such as English language learners, racial minorities, and low-income students (Klein, 2015a). Additionally, under NCLB, teachers must be "highly qualified" and highly qualified teachers should be evenly distributed among state schools.[6] Next, the Act set a goal to have all students testing at the 'proficient level' within 12 years from the end of the 2001–02 school year, allowing each state to define 'proficient' individually (NCLB, 2002). The second goal of NCLB is to reduce high school dropout rates by providing grants to schools to invest in school dropout prevention and re-entry programs (NCLB, 2002). Lastly, yearly progress is tracked, and if a school misses its adequate yearly progress targets for two or more years, then the school faces sanctions. The sanctions include allowing students to transfer to a higher-quality school, offering free tutoring to students, school shut down or forced change to a charter school, and/or setting aside a portion of federal Title I dollars for tutoring and school choice (borrowed heavily from Klein, 2015a).

The strength of school accountability, even post-NCLB, varied across states. First, each state has the freedom to design its own standardized accountability exam, allowing the difficulty of the exam to be determined by the state. The threshold for a proficient score is also determined at the state level. Next, the grades in which students are tested vary across states, and some states test students more than one time in each of the required testing periods. Moreover, although NCLB made all school accountability systems consequential by implementing sanctions, some states further emphasize the importance of the test by offering funding rewards to schools and/or teachers that have students who perform well on the exam. Lastly, some states implement promotion gates in the form of mandatory remedial classes, prohibited grade progression, or prohibited graduation if the state accountability exam is failed.

Prior to NCLB, not all states had school accountability policies, and those which did had complete autonomy in implementing the policy. Due to the lack of federal oversight on school accountability, state school accountability policies varied across the following aspects. First, states individually set performance goals for schools to achieve. Typically, these state-level targets varied across the following three dimensions: the expected level of student performance, the percentage of students that must meet these standards, and the length of time schools have to

---

[1] Deming et al. (2016) use variation in the pressure from accountability across cohorts to determine the effect of accountability on long term outcomes in Texas. They find cohorts that were in schools at risk of receiving a low-performance rating are more likely to graduate from high school and are more likely to attend and graduate from a four-year college. On the other hand, low-scoring students in cohorts at schools that were close to receiving a high-performance rating experienced decreased educational attainment.

[2] Carnoy and Loeb (2002) study school accountability for all states. However, their results for high school progression from tenth to twelfth grade are for students who were only exposed to school accountability for a few years and, similar to Deming et al. (2016), their empirical strategy compares students exposed to differing strengths of school accountability.

[3] Although the definition I use distinguishes between Title I and poorly ranked schools for funding, states without school accountability generally do not separately identify non-Title I schools for assistance.

[4] I distinguish between states that have report card accountability and consequential accountability because consequential accountability has a stronger bite in influencing student outcomes.

[5] Dee and Jacob (2011) demonstrate that consequential accountability states prior to NCLB had accountability policies that were comparable to NCLB, but there is no such evidence for report card accountability. Therefore, removing report card accountability states offers a cleaner identification because it is not comparing weaker treatments to stronger treatments or no treatment. See Dee and Jacob (2011) for more detail.

[6] "Highly qualified" means teachers must have substantial career experience, at least a Bachelor's degree, or for secondary schools, have at least a Bachelor's degree in the subject that they are teaching (NCLB 2002).

Table 1
Accountability adoption dates.

State	H&R Date	H&R Strength	D&J Date	State	H&R Date	H&R Strength	D&J Date
Illinois	.	.	1992	South Carolina	1999	Consequential	1999
Wisconsin	1993	Consequential	1993	Vermont	1999	Consequential	1999
Texas	1994	Consequential	1994	Tennessee	1996	Consequential	2000
Indiana	1995	Report Card	1995	Georgia	2000	Consequential	2000
Kansas	1995	Report Card	1995	Oregon	2000	Consequential	2000
Kentucky	1995	Consequential	1995	Mississippi	1994	Report Card	2002
North Carolina	1993	Consequential	1996	Minnesota	1996	Report Card	2002
Nevada	1996	Consequential	1996	Missouri	1997	Report Card	2002
Oklahoma	1996	Consequential	1996	Montana	1998	Report Card	2002
Alabama	1997	Consequential	1997	Washington	1998	Report Card	2002
Rhode Island	1997	Consequential	1997	Maine	1999	Report Card	2002
West Virginia	1997	Consequential	1997	Wyoming	1999	Report Card	2002
Delaware	1998	Consequential	1998	Arizona	2000	Report Card	2002
Massachusetts	1998	Consequential	1998	Hawaii	2001	Report Card	2002
Michigan	1998	Consequential	1998	Nebraska	2001	Report Card	2002
New York	1998	Consequential	1998	Colorado	2002	Report Card	2002
Virginia	1998	Consequential	1998	North Dakota	2002	Consequential	2002
New Mexico	2002	Consequential	1998	Utah	2002	Consequential	2002
Alaska	.	.	1998	Idaho	.	.	2002
Connecticut	1993	Consequential	1999	Iowa	.	.	2002
Arkansas	1999	Consequential	1999	New Hampshire	.	.	2002
California	1999	Consequential	1999	New Jersey	.	.	2002
Florida	1999	Consequential	1999	Ohio	.	.	2002
Louisiana	1999	Consequential	1999	Pennsylvania	.	.	2002
Maryland	1999	Consequential	1999	South Dakota	.	.	2002

Notes: The date school accountability was implemented in each state according to Hanesheuk and Raymond (H&R) (2005) and Dee ad Jacob (D&J) (2011) along with the accountability policy strength determined by Hanesheuk and Raymond. If a state's adoption date is 2002, then it did not have any accountability system prior to the NCLB Act. All states that have accountability implemented for Dee and Jacob are consequential.

meet their goals (Goertz & Duffy, 2001). Most states aimed to have a certain percentage of the student population reach 'proficient' on the state standardized exam.[7] The length of time a state allotted for the completion of the performance goal depended on the ambitiousness of the goal. Some states set smaller goals and a shorter time frame and incrementally increased the standards required to meet the goal, while other states started off with an aggressive goal and a longer time frame to complete it.

States also varied in how they defined progress towards completing these performance targets. Early school accountability policies generally used at least one of the following approaches to evaluate adequate yearly progress: (1) meet an absolute target, (2) make relative growth, and (3) narrow the achievement gap (Goertz & Duffy, 2001). Most states used an absolute target, where there is a specific performance threshold required in order to make satisfactory progress (Goertz & Duffy, 2001). The next most popular choice was a combination of relative growth and absolute target, followed by relative growth alone.[8] Lastly, some states included narrowing the achievement gap as part of their adequate yearly progress measure (Goertz & Duffy, 2001). Narrowing the achievement gap required schools to reduce the percentage of total students scoring at the lowest performance levels.

The last aspect for which there is variation in accountability across states is in the consequences of not meeting performance goals. When progress was not made and the performance goals were left unmet, most states implemented consequences at the school level directly. The consequences ranged from the mandatory development of school improvement plans to state takeover of the school or school closure (Goertz & Duffy, 2001). A few states also enacted promotion gates in

order to motivate students to exert sufficient effort on the accountability standardized exam. These promotion gates prevented students from progressing to the next grade if they did not meet the state or district performance standards. Additionally, school accountability could include school-level rewards, either monetary or non-monetary, for meeting the performance goal or demonstrating improvement towards reaching the goal.

In 2008, states began requesting some NCLB requirements be waived. States could receive waivers to opt out of particular provisions of the NCLB requirements, such as for the definition of low achieving schools. To apply for a waiver, states had to submit a proposal outlining how the waiver would increase the quality of instruction and improve achievement (Doan, 2008).[9] The goal of the waivers was to enhance achievement in an attainable way. In 2011, the waiver process became more formalized through the "Elementary and Secondary Education Act Flexibility waiver program" (Levesque, 2015). In November 2011, 11 states applied for the first set of waivers through the flexibility program, and all waivers were granted. By 2015, 42 states had waivers (Klein, 2015b).

The changes made to the bulk of waivers granted after 2011 revolved around (1) college- and career-readiness expectations, (2) state-developed differentiated recognition, accountability, and support, and (3) support of effective instruction.[10] Similar to accountability policies in general, the waivers have broad goals and significant flexibility in individual state freedom. Additionally, states have to meet certain requirements to qualify for a waiver which are similarly vague and flexible.[11] Therefore, the waivers should not be dramatically changing the general policy of school accountability and certainly not in a systematic way, but they will be part of the treatment effect.

Although school accountability varied across many aspects, such as

---

[7] The definition of 'proficient' and type of standardized exam varies by state. For example, most states use criterion-referenced assessments to judge student performance for accountability purposes, although some states use norm-referenced exams and other states use a combination of both (Goertz and Duffy 2001).

[8] The relative growth measure bases the future performance level on each school's past performance level and typically reflects the school's distance from the state goals (Goertz and Duffy 2001).

[9] The waivers could be granted for up to four years, during which states had to submit reports on their progress to Congress (Doan 2008).

[10] Ayers et al. (2012) Appendix A lists the flexibility the waivers would allow for.

[11] The reforms are described in Ayers et al. (2012) Appendix B.

through the way success was defined and measured and what the consequences for inadequacy were, it also shared common elements. All school accountability policies include standardized student-based assessments, proficiency standards, school-level performance reporting, and in most cases, consequences for low performance. The commonalities among state accountability policies pre- and post-NCLB allow accountability policies to be comparable across states and over time.

3. Conceptual framework

School accountability can affect educational attainment through various avenues and the direction of its long-run impact on educational attainment is ambiguous. School accountability can be expected to improve graduation rates in four ways. First, school accountability reduces the asymmetric information between the school and the parent by offering parents a measure of school quality. Informing parents about school quality will help them make more optimal decisions regarding school choice (Schellenberg & Walters, 2020; Cullen, Jacob & Levitt, 2005).

Second, in order to attract high-achieving students, public schools should be motivated to improve their accountability rating, and thus their overall quality (Dinerstein & Smith, 2021). Better school-student match and increased competition among public schools should improve student outcomes in the short run and potentially, in the long run. Third, some states implemented promotion gates along with the conditions of the accountability system, which prevent students from moving to the next grade if they don't meet the proficiency standard for the accountability test. Promotion gates may incentivize students or parents to put effort into ensuring the exam is passed because they do not want to be held back or they do not want their children to be held back.

Lastly, school accountability imposes explicit goals upon schools to ensure students meet specific learning and testing standards. Schools should be motivated to meet these goals as school accountability traditionally offers rewards (punishments) for schools that meet (don't meet) the performance goals.[12] Testing standards aim to directly improve student knowledge and test scores in the short run and hence may also lead to long-run effects.

On the other hand, school accountability could have adverse effects on student outcomes. First, school accountability increased the frequency of standardized testing. Increased test taking may discourage students who feel the curriculum is detached from their future careers, as test-taking ability is not a skill that has a direct relation to the skills required in the labor market. Second, promotion gates can directly cause declines in graduation rates by preventing students from graduating, and they can indirectly reduce students' motivation in school, leading to higher dropout rates.[13] Third, as has been shown in previous literature, teachers respond to these high-stakes tests by either teaching to the test or teaching to the median student (Krieg, 2008; Neal & Schanzenbach, 2010). When teachers teach to the median student, students in the tails of the distribution may be left behind.

Lastly, accountability tests are high-stakes tests because the schools can either receive funding if they successfully meet the testing standards or be punished if they fail to meet the standards. As a result, schools may create remedial courses to provide additional help for students who perform poorly on the accountability test. If this is the case, then the school would be grouping all low-performing students in the same course. Although these students may then pass the accountability test due to the extensive preparation, more exposure to the less academically inclined peers in their class may have a negative effect on other outcomes, such as graduation rates or college attendance.

---

[12] For states with report card accountability, the threat of a public rating should also give the school sufficient motivation to meet the accountability goals.
[13] Promotion gates can occur in multiple grades and which grade(s) they occur in depends on the state.

4. Data

I use the American Community Survey (ACS) for the years 2008 to 2016 (Ruggles et al., 2020). The ACS replaced the long-form of the decennial census as the key source of information about the American population and housing characteristics. The survey contains cross-sectional information about individuals' and households' education, ethnicity, labor-force participation, income assistance, immigration, and occupations, among others. The data consists of 36 million observations before narrowing the sample to the relevant population.

Importantly, the ACS is a large dataset that post-2008 distinguishes between receiving a GED and a traditional high school diploma. As this paper relies on high school graduation status for people who have already graduated, this is a significant feature of using the ACS. Additionally, the ACS records state of birth, but not the state where a person attended school. I, therefore, assign individuals to treatment based on the state in which they were born.[14] While the ACS has parental characteristics, such as mother's education, these characteristics are only available for individuals who are still residing in the household with their parents. Since my paper studies individuals after their school-going age, most individuals will no longer reside with their parents and this information is not available.

To estimate the effect of school accountability on student educational attainment, I compare individuals who were in school when the policy passed to students who were already out of school. Specifically, individuals who were at most 17 years old during or after the policy passed in their state are considered treated, and students who were 18 or older when the policy passed are considered untreated. I define individuals to be exposed to the policy until 17 years old as I can distinguish between GED recipients and traditional high school graduates. Therefore, if an individual is 17 years old when the policy passed in their state, then they will have been exposed to the policy for one year.

The analysis is restricted to individuals who are at least 20 years old in the survey year.[15] I also limit my sample to individuals who are at most 30 years old the year the policy passed in their state as older individuals are less similar to school-age individuals. After these restrictions, my sample size reduces to 6.1 million observations. Lastly, for my primary analysis, I exclude any state that at any point had report card accountability prior to NCLB. Unlike the consequential accountability policies of the 1990s, report card accountability has not been shown to be similar to NCLB (Dee & Jacob, 2011). As a result, the treatment for these students changes over my sample period. To avoid comparing individuals with different treatments effects, I omit these states from my main analysis.

The summary statistics in Table 2 show that states which adopted school accountability policies earlier were similar to late adopters in most respects; individuals in early-adoption states have higher incomes, are more likely to have a Bachelor's degree, and are older. States that adopted report card accountability and those that adopted consequential accountability policies before NCLB are also similar across characteristics. Individuals in states which adopted consequential accountability are slightly older and have slightly higher total incomes, similar to the early versus late adopters. These differences can be explained by the mechanical relationship between age, wealth,

---

[14] If individuals went to school in a different state from where they were born then I will be misclassifying their exposure to school accountability, which would attenuate my estimates. I have performed two robustness tests to check for this. First, I control for current state of residence in my analysis and demonstrate my results are largely unchanged in the Online Appendix Table A4. Second, I investigate if moving predicts exposure to school accountability; I find moving does not significantly predict exposure, shown in Online Appendix Table A5.
[15] Eighteen and nineteen-year-olds are excluded from the analysis because they have atypically low graduation rates.

Table 2
Summary statistics.

	Full Sample	Report Card Accountability (before NCLB)	Consequential Accountability (before NCLB)	Early Adopt ($<=$ 1998)	Late Adopt ($>$ 1998)
At least HS	0.853	0.870	0.864	0.870	0.841
	(0.354)	(0.336)	(0.343)	(0.336)	(0.366)
At least BA	0.293	0.269	0.289	0.304	0.285
	(0.455)	(0.444)	(0.453)	(0.460)	(0.451)
Age	29.25	27.66	29.86	30.72	28.20
	(6.17)	(4.98)	(6.62)	(6.95)	(5.29)
Male	0.502	0.503	0.502	0.501	0.503
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Total Income	27,346	25,357	28,353	29,981	25,443
	(32,335)	(28,391)	(33,392)	(34,727)	(30,349)
Employed	0.718	0.733	0.719	0.733	0.708
	(0.450)	(0.442)	(0.449)	(0.443)	(0.455)
White	0.740	0.797	0.768	0.802	0.695
	(0.439)	(0.402)	(0.422)	(0.399)	(0.460)
Black	0.118	0.082	0.135	0.124	0.113
	(0.322)	(0.274)	(0.342)	(0.330)	(0.317)
Hispanic	0.167	0.071	0.135	0.112	0.207
	(0.373)	(0.257)	(0.342)	(0.315)	(0.405)
Mother HS	0.534	0.480	0.529	0.527	0.539
	(0.499)	(0.500)	(0.499)	(0.499)	(0.498)
N	6135,025	469,455	4050,774	2573,410	3561,615

Notes: Summary statistics for individuals who are no younger than 25 years old in the policy year and no older than 60 in the survey year (the widest sample used in the analysis). Columns 2 and 3 do not sum to the full sample because they only include states who had accountability prior to NCLB. Income is inflation adjusted to 2010 dollars. Standard deviations are in parenthesis.

employment, and education.

School accountability policies were implemented between 1992 and 2002. After 2002 and the implementation of the NCLB Act, all states had school accountability. The school accountability implementation dates were collected from Dee and Jacob (2011). Dee and Jacob (2011) used Hanushek and Raymond's (2005) school accountability implementation dates as a base and did a comprehensive check of these dates, finding only a few discrepancies.[16] The states' policy implementation dates are displayed in Table 1.

I use the Common Core of Data (CCD) Local Education Agency (School District) Finance Survey for the fiscal years 1990 to 2016 to investigate potential mechanisms behind school accountability. The CCD Finance Survey is a survey that reports revenue and expenditure data for the universe of school districts in the United States.[17] All school districts providing public education from pre-kindergarten to twelfth grade are included. Data on the number of students eligible for special education at the school district level are from CCD's Local Education Agency (School District) Universe Survey Data from 1990 to 2019. The number of students enrolled in language assistance programs at the school district level are also from CCD's Local Education Agency (School District) Universe Survey Data and are available starting from the 1998–9 school year.

5. Methodology

This paper uses a dynamic differences-in-differences methodology with an intensity measure to estimate the long-run effects of school accountability. The main source of exogenous variation is the adoption of school accountability policies across states. The second source of variation is an individual's exposure to school accountability as a result of their age in the policy year. Fig. 1 is a stylized depiction of what a positive effect of accountability on student educational attainment could look like. The x-axis is an individual's age in the year the policy went into effect. All students who were out of school (18 years old or older) when the policy passed in their state make up my control group as they

were not exposed to school accountability. If an individual is six or younger, then they are fully exposed to school accountability. Fig. 1 displays no effect of the policy for the unexposed group, represented by zeros on the y-axis; increasing educational attainment among individuals who are partially exposed to the policy; and flattened (or increasing) educational attainment after individuals have reached full exposure to the policy.[18] After 12 years of exposure there may be a trend of continued improvement in educational attainment if teachers are learning how to effectively teach under school accountability or if later policies (such as NCLB) are more effective. The main specification for the event study plot is as follows,

$$y_{isbt} = \alpha + \sum_{a=-4}^{25} \beta d_{sba} + \lambda_s + \gamma X_i + \omega_b + \Phi_t + \epsilon_{isbt} \qquad (1)$$

where y_{isbt} is either high school graduation, college attendance, or college graduation[19] for person i born in state s with current age b in survey year t, α is a constant, λ_s is a birth-state fixed effect, d_{sba} is a dummy variable equal to one if an individual currently age b, born in state s, would be age a in the policy year in that state, X_i is a vector of covariates (race and sex), and ϵ_{isbt} is a random error term. The variables ω_b and Φ_t are current-age and survey-year fixed effects respectively. Survey-year and current age fixed effects are included to account for any current or past labor market conditions that could affect high school drop out, college enrollment, or college graduation.[20] Since state characteristics from one year to the next are correlated, robust standard errors are clustered at the state level.

---

[16] For more information on the discrepancies see Dee and Jacob (2011) Appendix B.

[17] For the years 1990-1, 1992-3, and 1993-4, a sample of school districts was collected.
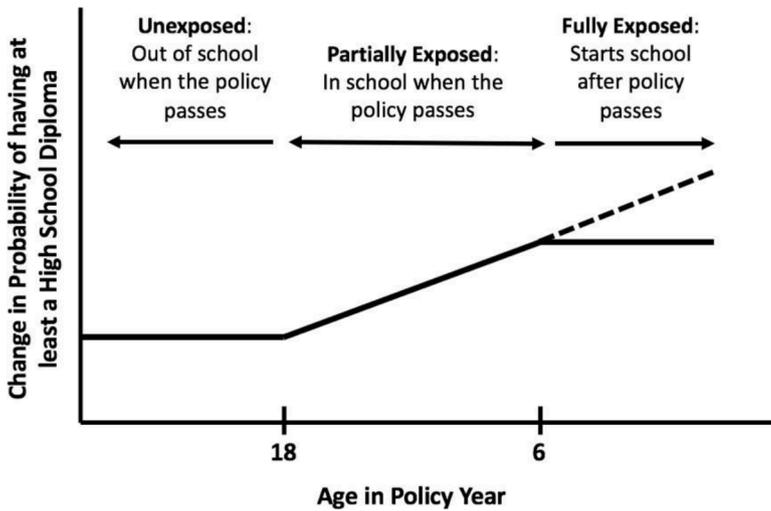
[18] This figure is designed under the assumption that accountability has a positive causal effect on educational attainment. The figure concept is based on Jackson et al. (2015).

[19] Online Appendix Table A3 displays the results for receiving an Associate's (AA) Degree. The results are largely similar to the enrollment estimates.

[20] Conceptually, an individual's high school graduation status cannot change after age 20 or 21, so there should be no concern that age effects or changes in the macro-environment (such as a recession) would cause a 25 year old's high school graduation rate to change. Since survey-year and current age fixed effects are theoretically irrelevant for this outcome, I show in the Online Appendix Figure A1 and Table A1 the pre-trends and regression estimates for high school graduation without survey-year and current age fixed effects. The results are consistent.

Fig. 1. Model of effect of school accountability on outcomes.
Notes: A stylized depiction of how a positive effect of school accountability on student educational attainment could look. The x-axis is an individual's age in the policy year. If an individual is 18 or older, they are not affected by accountability and if an individual is 6 or younger, they are fully affected by accountability. There is no effect of the policy for the unexposed group, represented by zeros on the y-axis; increasing educational attainment among individuals who are partially exposed to the policy; and flattened (or increasing) educational attainment after individuals have reached full exposure to the policy. Figure heavily borrowed from Jackson et al. (2015).

Table 3
Age restrictions.

Outcome	Min Survey Year Age	Max Survey Year Age	Max Policy Year Age
High School Graduation	20	60	25
College Attendance	24	40	27
College Graduation	27	40	30

Notes: Age restrictions imposed for the main specification by outcome.

In order to ensure the comparison group is similar to the treatment group, I impose age restrictions in each regression. I established the age restrictions by determining at what age average individuals' educational attainment levels off; found in Fig. A2 in the online Appendix. From these graphs, I impose an age cap of 60 years old in the survey year for high school graduation and a cap of 40 years old in the survey year for college attendance and college graduation. Each regression's minimum age restriction varies depending on the outcome variable. For high school graduation, I include individuals who are at least 20 years old in the survey year, college attendance includes individuals who are at least 24 in the survey year, and college graduation includes individuals who are at least 27 years old in the survey year.

As mentioned above, I only include individuals who are reasonably close to school-going age in the policy year. For high school graduation, individuals who are at most 25 years old in the policy year are included; for college attendance, individuals who are at least 27 years old in the policy year are included; for college graduation, individuals who are at least 30 years old in the policy year are included. Table 3 lists the age restrictions in a concise way. I provide a robustness check of these age restrictions in Figs. A3, A4, and A5 in the online Appendix. Table A2 in the online Appendix shows all three measures of treatment using the Bachelor's degree age restrictions to demonstrate the effect is not being driven by the changing sample.[21]

In the main analysis, I impose two restrictions on the policy year age dummies from Eq. (1). First, individuals who are six years old or younger in the policy year are all fully exposed to school accountability and are assigned 12 years of exposure. Second, individuals who are at least 18 years old at the time of the policy are assigned zero exposure to the

policy. After these restrictions, Eq. (1) becomes

$$y_{isbt} = \alpha + \sum_{e=}^{12} \rho I_{sbe} + \lambda_s + \psi X_i + \omega_b + \Phi_t + \mu_{isbt} \qquad (2)$$

where I_{bse} is a dummy variable equal to one if an individual currently age b, in state s is exposed to e years of school accountability. The intensity measure is restricted to have a minimum value of 0, no exposure to school accountability, and a maximum value of 12, full exposure to school accountability.[22] Robust standard errors are again clustered at the state level.

The identification strategy relies on two main assumptions. First, there is no policy endogeneity. That is, school accountability could not have been adopted in response to a dip in high school graduation rates, college attendance, or college graduation rates. This would primarily be of concern for the states that adopted accountability policies prior to NCLB. If this were the case, it will be seen in the pre-trends on the event study graphs.

The second assumption is that there are no contemporaneous policies or simultaneous treatment effects. The existence of contemporaneous policies could be a concern in this case because there were other educational policies passed around the same time period as when school accountability was being implemented. For instance, the Improving Americas Schools Act of 1994, created major changes in Title I and granted additional funds to schools with large concentrations of low-income children. Additionally, many individual states implemented spending reforms during this period. In a difference-in-differences model, this would be a concern if a secondary policy happened at the same time as the policy of interest. I perform a robustness check for the inclusion of spending reforms to test for the presence of contemporaneous expenditure reforms.

6. Results

In this section, I demonstrate the validity of the research design,

---

[21] The main specification does not use the Bachelor's degree sample because it would exclude many fully treated individuals who have graduated high school or could have attended college.

---

[22] In principle, a more restrictive model has the potential to increase power if the restricted specifications are indistinguishable from the more flexible specification. Therefore, I compare the treatment effects from the fully flexible model in Equation (2) to a linear model, and a binary specification for whether a student was exposed to school accountability for any length of time. I use an F-test to compare the fits across models. The estimates from the different specifications for high school graduation are statistically different from each other. Hence, I reject that the other specifications are sufficient and use the fully flexible model as my main specification.

discuss my main results, and provide evidence that the results are consistent with a balanced panel. I find that full exposure to school accountability increases high school graduation rates by an average of 2.3 percentage points but has no significant effect on college attendance or college graduation by the minimum age of 27.

6.1. Identification validity check

Fig. 2 displays the event study graph for high school graduation that comes from Eq. (1). The first vertical line represents the time treatment began and the second vertical line represents the time individuals become exposed to accountability for twelve years of school. The omitted group is individuals who were 18 years old in the policy year. Following from Fig. 1, if the estimates are zero before the first vertical line, then it is likely the treated and untreated individuals were trending similarly prior to the policy implementation. The points before the first vertical line in Fig. 2 are statistically indistinguishable from zero, suggesting states were trending similarly before accountability was enacted.

As expected, high school graduation increases with exposure to school accountability, demonstrating school accountability has a positive effect on high school graduation. The effects stay rather strong after 12 years of exposure, implying there may be learning by teachers or schools, or adjustments by districts and states on how to teach effectively under school accountability. There is evidence that schools were adjusting their school accountability policies even after the policy implementation date. For example, many states were still developing the NCLB requirements of reporting separate proficiency rates for each student subgroup and testing at each school level years after the enactment of NCLB (Taylor et al., 2010).[23]

Fig. 3 shows the event study for college attendance. It is clear from Fig. 3 that states were trending similarly before school accountability was implemented and there is a positive but statistically insignificant effect of school accountability on average college attendance. Fig. 4 shows the event study for college graduation. Akin to the event study for college attendance, states were trending similarly before the policy passed, but there doesn't appear to be a large effect of school accountability on college graduation, if anything the effect appears to be negative.

6.2. Main results

Table 4 shows the regression estimates that follow from Eq. (2) which are comparable to the event study graphs in Figs. 2, 3, and 4.[24] Similar to the figures, one can see that the estimates for high school graduation grow larger and more significant as exposure to accountability increases. According to Table 4, being exposed to school accountability for 12 years leads to an average increase in the likelihood of graduating high school by 2.3 percentage points.

The results for college attendance are positive but insignificantly different from zero; the estimates for graduating from college with a Bachelor's degree by at least 27 years old are less consistent and statistically insignificant. Table 4 also demonstrates that the effects are

concentrated among those with more exposure to school accountability.[25] It is likely the earlier grades adjusted their teaching style or content to prepare the students for their first potentially high stakes standardized test.

A 2.3 percentage point increase in the high school graduation rate is economically significant and is evidence that school accountability has a long-term positive effect on student success. To put this figure into context, Loeb and Page (2000) find that increasing teacher wages by 10% reduces high school dropout by between 3 and 6 percent. I find school accountability increased high school graduation rates by about 2.7 percent (off a mean graduation rate of 85 percent for the full sample from Table 2). Hence, full exposure to school accountability has a comparable effect to a 10% increase in teacher wages on high school completion.

Relative to other studies of school accountability, the effect on high school graduation is slightly larger while the college educational attainment results are smaller (Deming et al., 2016). Deming et al. find students who face high-pressure school accountability are significantly more likely to attend and graduate from a four-year college, but students who face low-pressure school accountability are less likely to both attend and graduate from college. Because my specification includes both students who are exposed to high-pressure and low-pressure school accountability, it follows that the college results would fall below the estimates found by Deming et al. Further, the college estimates are consistent with the fact that previous studies have found test score improvements for students in the middle of the distribution but decreases in performance in the tails of the distribution (Neal & Schanzenbach, 2010). Students in the middle of the distribution may have been on the margin of graduating high school, but were never on the margin of going to college.

7. Heterogeneity

In this section, I examine heterogeneity in the effect of school accountability policies by student race and by the strength of school accountability. I find Hispanic students exhibit the largest gain in the likelihood of attending college while Black students have insignificant positive gains in high school graduation rates but may be less likely to graduate from college.

For the strength of school accountability, I define a student as being exposed to strong school accountability if they are exposed to both high school exit exams and school accountability, as states that implement high school exit exams are likely to have stronger accountability policies. I find exposure to both high school exit exams and school accountability leads to larger gains in high school graduation rates than school accountability alone.

7.1. Student race

One goal of school accountability is to reduce inequities in the American education system. Therefore, after the implementation of NCLB, schools were required to report race-specific test scores and meet race-specific performance targets. Thus, the effect of school accountability may differ by race depending on where Black, White, and Hispanic students on average fall in the distribution of test scores. As a result, I perform a heterogeneity analysis by race.

Table 5 shows the results of the heterogeneity analysis. The estimates for White students are smaller than the main estimates; White students are 1.5 percentage points more likely to graduate high school on average under school accountability. Black students are more likely to graduate

---

[23] According to a U.S. Department of Education report, "As of 2006-07, 37 states, the District of Columbia, and Puerto Rico, had developed entirely new tests or modified existing assessments in grades 3-8 to comply with NCLB. By 2006-07, all states had alternate assessments for students with disabilities, and one-third of states had developed entirely new alternate assessments based on alternate academic achievement standards." Further, states frequently adjusted their content standards for reading, math, or science; with 30 states updating them on a regular schedule as of 2006-07. Lastly, the report found the NCLB requirement to test once in each of the schooling levels, was one of the most substantive challenges states faced in the first three years of NCLB (Taylor et al. 2010).

[24] Recall, these results exclude states that ever had report card accountability prior to NCLB.

[25] Students with greater exposure to the policy are those who were exposed to school accountability from younger grades. For instance, someone exposed for 8 years would experience school accountability from 4th grade on, while someone with 12 years would experience accountability from kindergarten on.
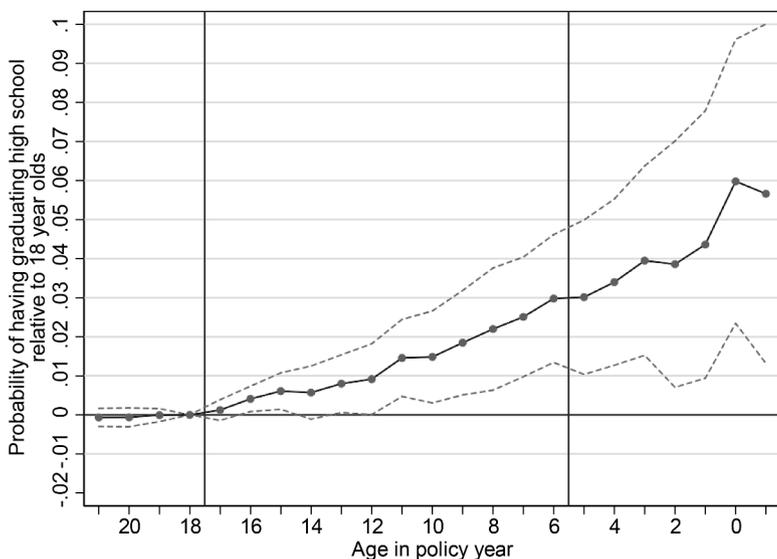
Fig. 2. Pre-trends graph for high school graduation.
Notes: Event study for high school graduation from Eq. (1). The first vertical line represents the time treatment began and the second vertical line represents when individuals are exposed to accountability for all twelve years of their schooling. The omitted group is individuals who were 18 in the policy year. There is no trend before the first vertical line, suggesting that until policy implementation, treated and untreated individuals experienced similar trends in high school graduation rates. States that ever had report card accountability are excluded from this analysis. Robust standard errors clustered at the birth-state level.
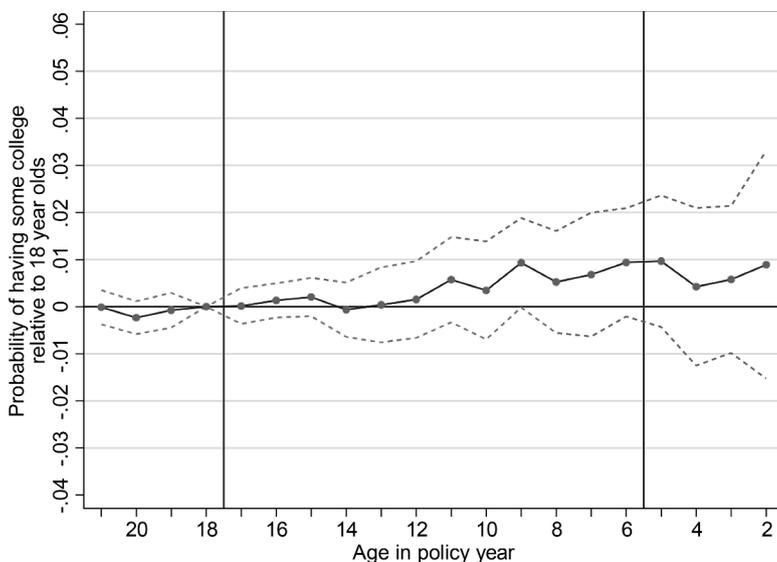
Fig. 3. Pre-trends graph for college attendance.
Notes: Event study for college attendance from Eq. (1). The first vertical line represents the time treatment began and the second vertical line represents when individuals are exposed to accountability for all twelve years of their schooling. The omitted group is individuals who were 18 in the policy year. There is no trend before the first vertical line, suggesting that until policy implementation, treated and untreated individuals experienced similar trends in college attendance rates. States that ever had report card accountability are excluded from this analysis. Robust standard errors clustered at the birth-state level.

high school by 3.5 percentage points and are less likely to attend and graduate from college with a Bachelor's degree, although these estimates are not significantly different from White students.

Hispanic students are the most positively affected. Hispanic students are 2.9 percentage points more likely to graduate high school and 4.4 percentage points more likely to graduate college with a Bachelor's degree by at least age 27, these estimates are not statistically different from White students. However, Hispanic students are 7.7 percentage points more likely to attend college and this is significant at the 1% level. These findings are in line with previous studies which have found heterogeneity in the effect of school accountability by race; typically finding positive effects for Hispanic students (Carnoy & Loeb, 2002; Dee & Jacob, 2011).

7.2. Strength of accountability

Some states implemented promotion gates in the form of exit exams with, or prior to, their school accountability policies. Exit exams are exams that students must pass in order to receive their high school diploma. Most states implemented exit exams in the 1980s (before consequential school accountability). But some states continued to

implement exit exam policies through the 1990s and 2000s.[26] The presence of exit exams might indicate that school accountability policies are stronger in these states. Alternatively, exit exams may have complementarities with school accountability as both emphasize testing.

Table 6 displays the results with an interaction for whether an individual was exposed to an exit exam before graduation.[27] Consistent with previous literature, exit exams alone may reduce high school graduation (Baker & Lang, 2013; Dee & Jacob, 2006). School accountability alone still has a positive effect of 0.8 percentage points on the likelihood of graduating high school. Since the effect of school accountability is positive even after controlling for exit exams, it is clear the effect is not due entirely to exit exams. Further, the magnitude is still economically relevant; Deming et al. (2016) found that the high school

---

[26] Exit exam dates are from Baker and Lang (2013) for 1980 to 2001 implementation and McIntosh (2012) for 2002 to 2016 implementation.

[27] Exit exam dates are defined at the individual level for the first cohort to graduate with an exit exam requirement. After a state passes an exit exam policy all individuals who were younger than 18 in that state have an indicator for being exposed to an exit exam.
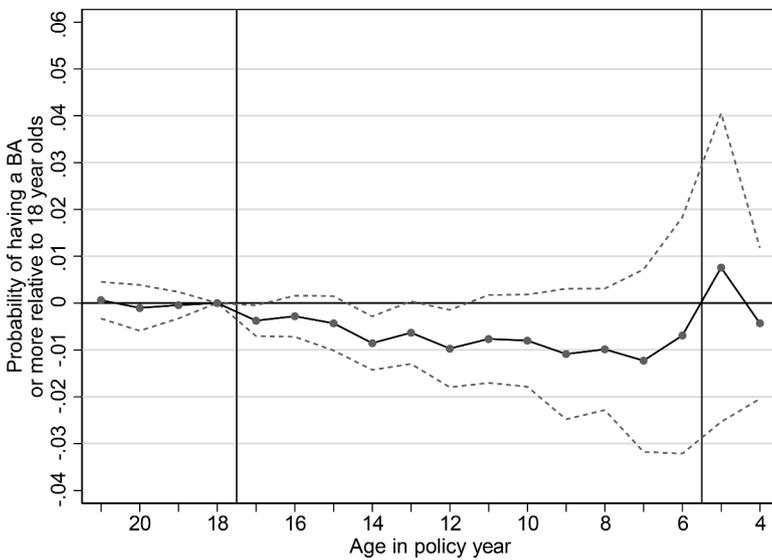
Fig. 4. Pre-trends graph for college graduation (BA Degree).
Notes: Event study for college graduation from Eq. (1). The first vertical line represents the time treatment began and the second vertical line represents when individuals are exposed to accountability for all twelve years of their schooling. The omitted group is individuals who were 18 in the policy year. There is no trend before the first vertical line, suggesting that until policy implementation, treated and untreated individuals experienced similar trends in college graduation rates. States that ever had report card accountability are excluded from this analysis. Robust standard errors clustered at the birth-state level.

Table 4
Effects by years of exposure.

Dependent Variable	HS Diploma	College Attendance	BA Degree
1 Year	−0.001	0.001	−0.002
	(0.003)	(0.002)	(0.002)
2 Years	0.001	0.003	−0.001
	(0.003)	(0.002)	(0.002)
3 Years	0.002	0.004	−0.002
	(0.004)	(0.003)	(0.003)
4 Years	0.001	0.001	−0.005*
	(0.004)	(0.003)	(0.003)
5 Years	0.003	0.002	−0.002
	(0.005)	(0.004)	(0.004)
6 Years	0.003	0.004	−0.005
	(0.004)	(0.004)	(0.004)
7 Years	0.008*	0.008**	−0.003
	(0.004)	(0.004)	(0.005)
8 Years	0.007*	0.006	−0.002
	(0.004)	(0.005)	(0.005)
9 Years	0.010**	0.012**	−0.005
	(0.005)	(0.005)	(0.007)
10 Years	0.013**	0.008	−0.003
	(0.005)	(0.005)	(0.006)
11 Years	0.015**	0.010	−0.005
	(0.006)	(0.007)	(0.010)
12 Years	0.023***	0.011	0.006
	(0.007)	(0.007)	(0.012)
Race, ethnicity, & gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Adj. R-squared	0.024	0.043	0.055
N	4,828,380	3,736,937	3,215,494

Notes: Regression estimates from Eq. (2) for high school graduation, college attendance, and graduating with a Bachelor's. Robust standard errors are clustered at the birth state level.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

Table 5
Heterogeneity by race.

Dependent Variable	High School	College Attend	BA Degree
Full Exposure (White)	0.015***	0.006	0.009
	(0.005)	(0.007)	(0.018)
Full Exposure × Hispanic	0.014	0.071***	0.035
	(0.023)	(0.018)	(0.021)
Full Exposure × Black	0.020	−0.012	−0.025
	(0.015)	(0.019)	(0.021)
Gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Adj. R-squared	0.026	0.043	0.051
N	4,578,887	3,557,267	3,072,506

Notes: Regression estimates for heterogeneity by race for full exposure (12 years) to school accountability using the model specification from Eq. (2). States that ever had report card accountability are excluded from this analysis. Robust standard errors are clustered at the birth state level. *$p < 0.05$, ** $p < 0.01$,.
*** $p < 0.001$.

Table 6
Interacting with exit exams.

Dependent Variable	High School	College Attend	BA Degree
Full Exposure (12 years)	0.008 (0.007)	0.006 (0.008)	0.013 (0.009)
Exit Exam	−0.009	−0.007*	0.002 (0.006)
	(0.008)	(0.004)	
Exit Exam × Full Exposure	0.023***	0.009*	−0.024***
	(0.007)	(0.006)	(0.008)
Race, ethnicity, & gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Adj. R-squared	0.024	0.043	0.055
N	4,828,380	3,645,737	3,141,613

Notes: Heterogeneity in accountability strength for full exposure (12 years) to school accountability using the model specification from Eq. (2). Exit exams are used as a proxy for strong school accountability. States that ever had report card accountability are excluded from this analysis. Robust standard errors are clustered at the birth state level.
* $p < 0.05$, ** $p < 0.01$,.
*** $p < 0.001$.

graduation rate increased by approximately 0.9 percentage points on average for students who faced high accountability pressure. The consistency of the high school graduation rate estimate demonstrates that school accountability has a positive effect on educational attainment on average and not only for students who face high accountability pressure.

Under the presence of exit exams, students with full exposure to

school accountability are 3.1 percentage points more likely to graduate high school, on average.[28] Additionally, there is a significant positive effect for college attendance and a negative effect for likelihood of graduating with a Bachelor's degree. These results suggest accountability is stronger in the presence of exit exams, which can be due to either complementarities or the fact that states which implement exit exams have more stringent accountability policies.

8. Mechanisms

School accountability is a nebulous policy in which individual states have significant control over its implementation, even in the post-NCLB era. As a result, it is important to investigate what might be driving the results reported in this paper. In this section, I examine several potential mechanisms of school accountability including increased tracking, school district expenditures, teacher qualifications, and promotion gates. I find that spending and teacher qualifications are not changing as a result of school accountability, but there is suggestive evidence that schools are increasing the number of students classified as needing special education.

8.1. Tracking

Schools may respond to school accountability by increasing tracking in the school, such as by creating remedial classes. Although I am unable to directly test for tracking in schools and the introduction of remedial classes, I observe the proportion of students who are eligible for special education and students enrolled in language assistance programs. A higher proportion of students being classified as eligible for special education or language assistance programs after the adoption of school accountability, may be suggestive evidence that schools are increasing tracking as a result of school accountability.

Fig. 5 shows an event study graph of the proportion of students eligible for special education relative to the policy year. Special education eligibility is defined as students in the Individualized Education Program, which is a written document that is developed for each public-school child who is eligible for special education. The proportion of students eligible for special education seems to have increased slightly after school accountability, but the estimates are too noisy to determine if it's significantly different from zero. This offers suggestive evidence that schools may be more stringent about identifying students eligible for special education, a finding consistent with Cullen and Reback (2006).

One explanation for the positive impact on Hispanic students' educational attainment, could be if schools increased their support of students in language assistance programs. Examples of language assistance programs include English as a Second Language, High Intensity Language Training, and bilingual education. Schools may increase their support these students by increasing enrollment in language assistance programs or offering additional classes to help these students learn English. Fig. 6 displays the proportion of students in language assistance programs relative to the policy year.[29] Although the estimates are noisy, after the implementation of school accountability there is no change in the trend of the proportion of students enrolled in language assistance programs; hence school accountability does not seem to affect the proportion of students in language assistance programs. Overall, the lack of an effect on the proportion of students in language assistance programs

suggests schools are not implementing stricter tracking for these students to cope with school accountability.

8.2. School expenditures

After school accountability, states or districts could have decided to increase funding or the way that they distribute funding to schools. Fig. 7 shows three figures of different school expenditure measures. All graphs are relative to the year the policy was implemented and are in log real values. Panel A shows the per-pupil total expenditures of school districts. There appears to be an increasing trend in total expenditures over time, but it is unrelated to the implementation of school accountability. Panel B shows the per-pupil total current expenditures on elementary and secondary school miscellaneous. Again, there is a trend of increasing expenditures on elementary and secondary school, but it is not related to the enactment of school accountability. Panel C shows the same graph for per-pupil total current expenditures on instruction; it has an increasing trend similar to Panel A, suggesting school accountability did not lead to more spending on instruction. All three spending graphs show a consistent increase in spending over time, even in the period before school accountability policies. If this is expected to be driving the results, then the pre-period for the accountability figures would also show an increasing trend, but the pre-periods for student education attainment hovers around zero.
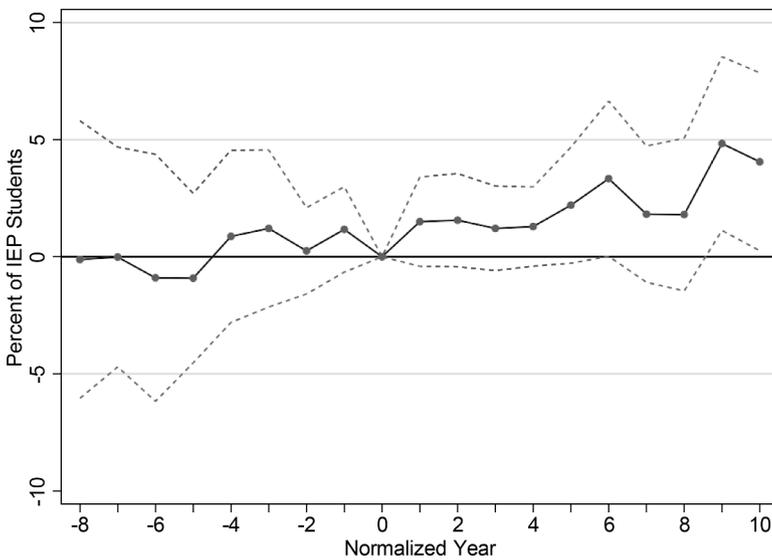
Total school expenditures may not have changed after school accountability was passed if states or districts decided to shift the funding from high revenue areas to low revenue areas. For example, a high-revenue district may have received less revenue from the state for education spending, but it could have increased its local revenue to counterbalance the lower state revenue, leading to no change in total expenditures. Fig. 8 shows the per-pupil log real total revenue for districts at the state and local levels. Revenue in this case is money that is received by the district from the state or local area to distribute to schools. Districts are binned into low-revenue or high-revenue districts before an accountability policy is passed in that state using the mean revenue for a particular year. If states or districts were giving low revenue areas more money in the post-school accountability period, then the dashed line should increase after the policy is implemented in Fig. 8. Panel A shows that the state possibly gave more revenue to high-revenue districts after school accountability. Panel B shows that local revenue for low-revenue districts may have decreased post-school accountability. Therefore, Fig. 8 provides suggestive evidence that the improved high school graduation rates are not resulting from a redistribution of spending from the state government. It may even be the case that low-revenue local areas are spending less on education.

8.3. Teacher characteristics

School accountability applies a lot of pressure on schools to ensure teachers are performing well. As a result, after school accountability, schools may be willing to pay a higher wage to more qualified teachers or increase the wage of teachers who demonstrate they are effective teachers under the scrutiny of school accountability.

Fig. 9 plots the log real wage of teachers. Panel A is the plot for all teachers; there is no significant change in the log real wage for teachers in general after school accountability was adopted. There is an increasing trend prior to the implementation of school accountability and after school accountability is implemented, wages flatten out. This suggests any increase in high school graduation is unlikely due to higher teacher wages after school accountability.

However, there could be a change for particular types of teachers if schools substitute between different types of teachers; e.g., schools can hire fewer or more assistant teachers because of school accountability and will adjust wages for these individuals as a result. Panels B and C show the log real wage of elementary and middle school teachers and secondary school teachers respectively; again, wage is not affected by

---

[28] The effect of school accountability with exit exams can be obtained by adding the estimate for full exposure and the estimate for the interaction term: $0.008 + 0.023$.

[29] Because enrollment in language assistance program data is only available for years 1999 and later, this analysis only includes states that implement school accountability in 1999 or after, ensuring included states have a pre-period.

Fig. 5. Proportion of students eligible for special education.
Notes: The proportion of students eligible for special education relative the policy year. Special education eligibility is defined as students in the Individualized Education Program, which is a written document that is developed for each public-school child who is eligible for special education. The regression that produced this plot includes dummy variables for years relative to the policy year and state and year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level. Data from the Common Core of Data Local Education Agency (School District) Universe Survey (years 1990 - 2019).
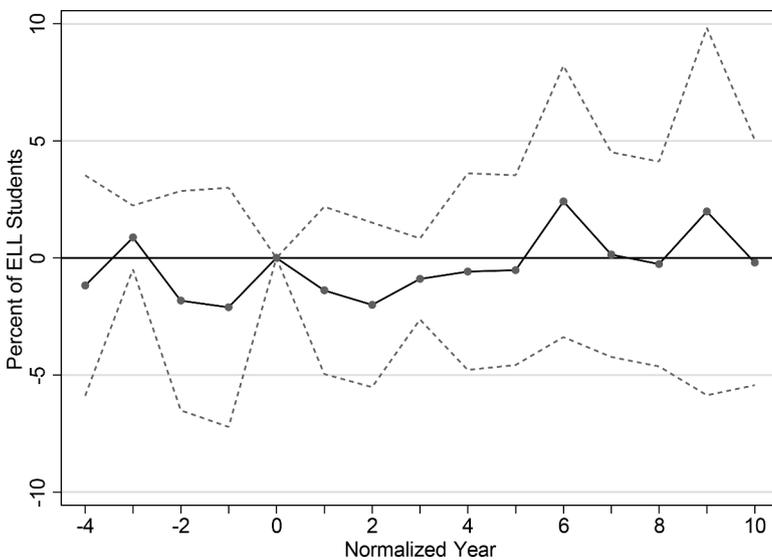
Fig. 6. Proportion of students in language assistance programs.
Notes: The proportion of students in the language assistance programs (e.g. English as a Second Language, High Intensity Language Training, bilingual education) relative to the policy year. The regressions that produced this plot include dummy variables for years relative to the policy year and state and year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level. Data from the Common Core of Data Local Education Agency (School District) Universe Survey (years 1999 - 2019).

school accountability. Panel D shows the log real wage of special education teachers and Panel E shows the log real wage of assistant teachers. Schools could consider both special education and assistant teachers to be good substitutes for full-time teachers. Although noisy, schools may be substituting away from assistant teachers as their log wage declines after the implementation of school accountability. On the other hand, there is no evidence that schools are using special education teachers to substitute for other teachers, as special education teachers' salaries do not change after school accountability.

Alternatively, schools may try to hire more qualified teachers. To investigate along this margin, I look at the types of degrees that teachers have relative to the policy year. Assuming graduate education is rewarded in the teacher labor market, ex-ante I would expect more teachers to have degrees above a traditional Bachelor's degree after the adoption of accountability. Fig. 10 shows the proportion of teachers with either a Bachelor's, Master's, Professional, or Doctoral degree. Fig. 10 suggests that more teachers have Bachelor's degrees than Master's degrees after school accountability was passed. However, from Fig. 10, it appears school teachers' educational qualifications were trending in this direction before school accountability was implemented. Consistent with the literature on the effects of NCLB on teacher

composition, schools do not appear to respond to school accountability by changing the composition of teachers that they hire (Hanushek & Rivkin, 2010).

8.4. What is left?

The mechanisms I explore in this paper are not exhaustive; there are many observable potential mechanisms that are not examined in this paper. For instance, previous literature and anecdotal evidence suggest that schools changed the format of their school days in order to improve their performance on school accountability exams. Schools may reduce the time students spend in elective classes and increase the amount of time students spend in core classes. It may also be the case that schools or teachers offer additional after-school support for students who are not excelling under school accountability. Furthermore, schools may adjust the curriculum they teach, and the new curriculum could help students develop skills that also improve educational attainment. In addition to the mechanisms I explore above, there are others that school accountability could be working through, and I leave those open to future research.
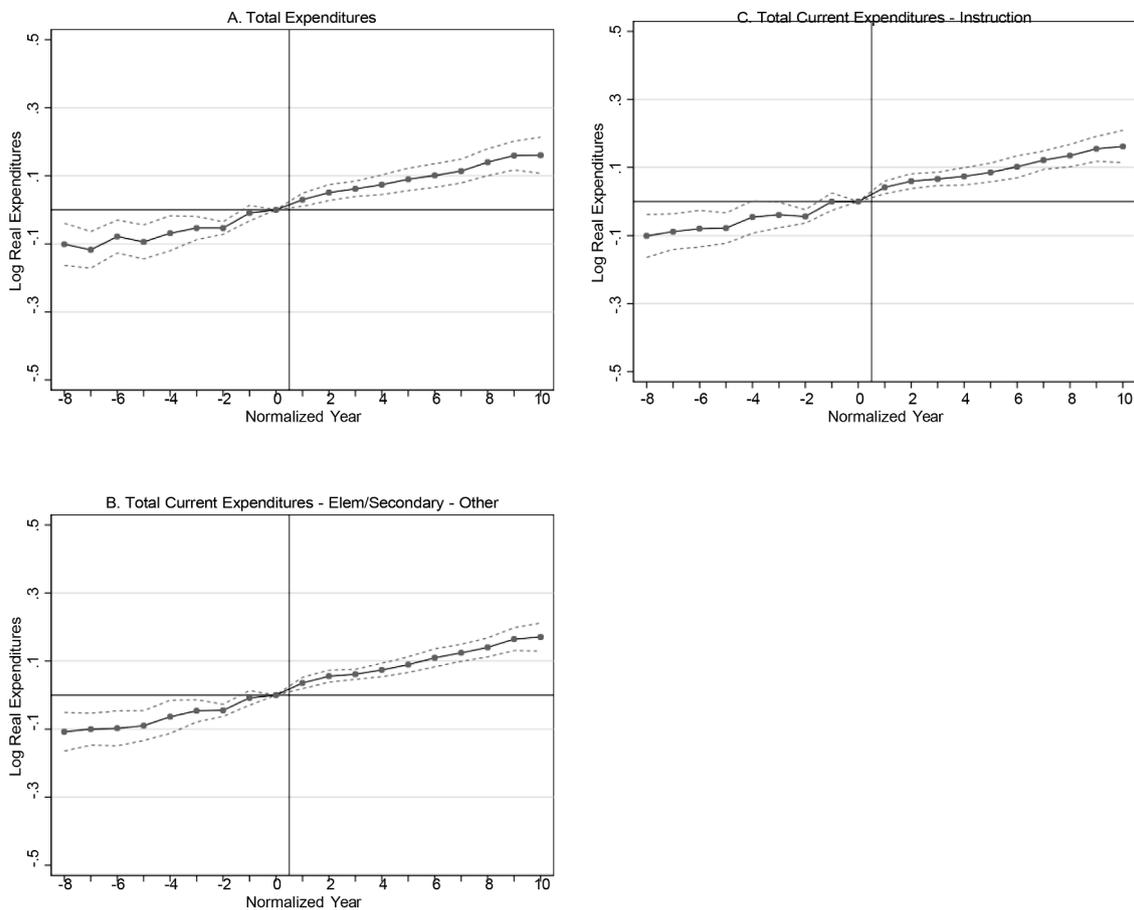
Fig. 7. School district expenditures.

Notes: The figure shows graphs for the log of real per pupil current expenditures relative to the policy year. The regressions that produced this plot include dummy variables for years relative to the policy year and state and year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level. Data from Common Core of Data School District Finance Survey (years 1990–2016).

9. Robustness

In order to test for the stability of the results to specification changes, I do three robustness tests: performing the main analysis using a balanced panel, including expenditure reforms, and including "report card" school accountability states. The estimates are consistent for all three robustness checks.

To begin, I discuss the results of my main specification with a balanced panel. Because school accountability was implemented by all states by 2002, the number of individuals who are fully exposed to school accountability (12 years) decreases as the age they were in the policy year decreases. That is, individuals who are younger in the policy year are also more likely to be younger in the survey year and are less likely to be of age to graduate high school. To clarify, the sample of fully treated is lower for later adopting states.[30] As a result, if younger individuals are more affected by school accountability due to teachers learning how to effectively teach under school accountability or if the later policies are more effective, then the estimates will be downward biased.

There are two ways of checking for the robustness of my estimates using a balanced panel. First, I can omit states that passed the policy in 1998 and earlier; this will enable me to have the same states over the full analysis. These estimates can be found in Panel A of Table 7. While the estimates are slightly smaller than the main results, supporting that NCLB may be a more effective policy in influencing high school graduation than the earlier accountability policies, they are qualitatively similar and have the same sign, reflecting the robustness of my results.

Second, I can include individuals only up to a certain age in the policy year. Individuals who are younger than 8 years old in the policy year have fewer observations in my data. As a result, I perform an analysis only including individuals who are at least 8 years old in the policy year. Panel B in Table 7 presents the results for this specification. Full exposure to the policy is unavailable for this group so I present the results for 10 years of exposure. The results are qualitatively similar to those in Panel A and the main results.

Next, using my main specification, I include controls for exposure to expenditure reforms that were passed between 1989 and 2010. The expenditure reforms included are from Supreme Court rulings on the constitutionality of school finance systems in reaction to the within-state differences in per-pupil spending due to the large portion of school funding coming from local property taxes. The reforms generally altered the parameters of spending formulas with the goal of reducing the unequal distribution of spending across low- and high-income students.[31]

---

[30] The states which first implemented school accountability in 2002 (NCLB only states) do not contribute to the college attendance or college graduation outcomes because these individuals are too young in the survey year to contribute. States that adopted accountability in 2002 are contributing to my estimates for high school graduation. I show the estimates for high school graduation of fully exposed individuals in Online Appendix Table A6. Exposure to NCLB accountability, shown in Online Appendix Table A7, yields larger effects, implying NCLB may be a more potent policy.

[31] For more information regarding the expenditure reforms and the specific dates see Appendix A in Jackson et al. (2015).
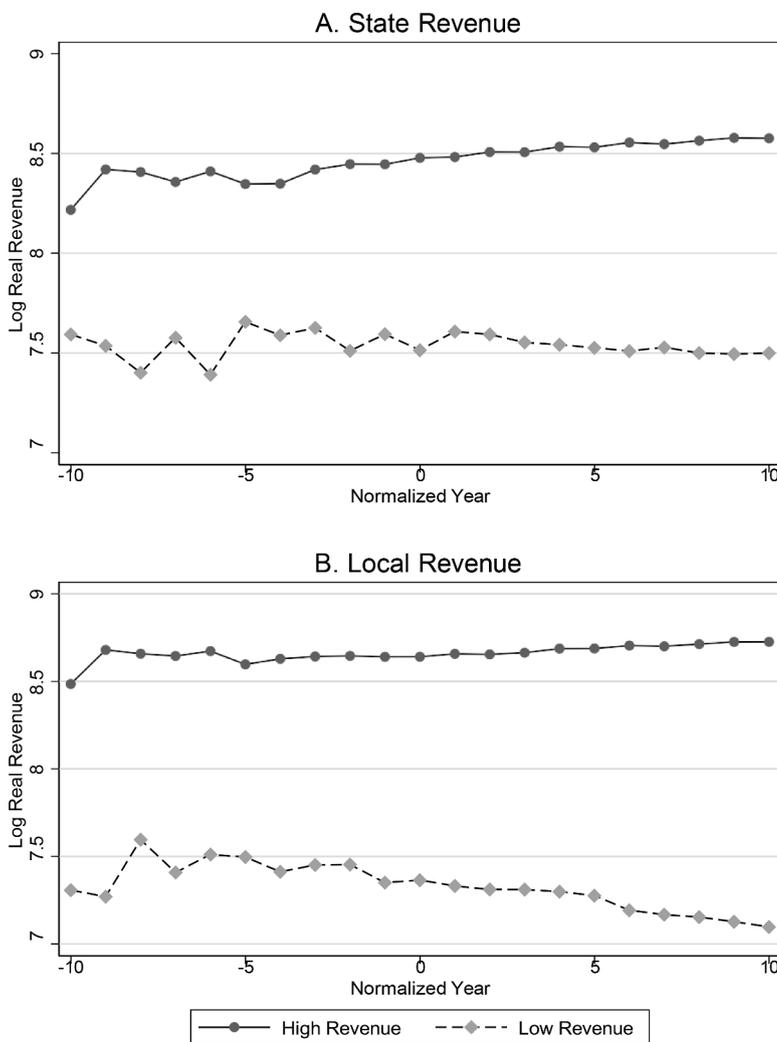
Fig. 8. School district expenditures by state and local revenue.
Notes: Graphs of the log of real per-pupil state and local revenue amounts received by the districts relative to the policy year. The regressions that produced this plot include dummy variables for years relative to the policy year and state and year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level. Data from Common Core of Data School District Finance Survey (years 1990–2016).

Table 8 shows full exposure to school accountability leads to an average increase in the likelihood of graduating high school by 2.1 percentage points. The estimate for high school graduation is smaller in magnitude but still positive and significant, suggesting the results are consistent even after controlling for school expenditure reforms that happened over the same time period as school accountability.

Lastly, the main specification excludes any state that had report card accountability implemented before NCLB. Report card accountability refers to accountability policies where school performance is made available to the public but there are no consequences, in the form of repercussions or additional funding, based on school performance. Report card accountability states are omitted from the main analysis to ensure consistency of treatment across all years. Table 9 demonstrates that including report card accountability states does not change the magnitude of the results.[32] Including the additional observations increases the power for college attendance and students are significantly more likely to attend college as a result of school accountability.

10. Conclusion and future work

Despite the importance of long-term outcomes in education, few

papers have been dedicated to investigating long-run effects. Using variation in the timing of policy adoption and exposure to school accountability based on an individual's age in the policy year, I estimate the effect of school accountability on high school graduation, college attendance, and college graduation. This paper builds on the existing literature by precisely estimating the long-run effects of varying levels of exposure to school accountability for all states.

The results show that exposure to 12 years of school accountability on average leads to an increase in the likelihood of graduating high school by 2.3 percentage points, and does not significantly affect college attendance or college graduation. The results are robust to a balanced panel analysis, different age restrictions, and the inclusion of expenditure reforms and report card accountability states.

Previous literature has found heterogeneous effects by race and because NCLB requires states to report proficiency rates for the school in general as well as for minority groups, I examine the effect of school accountability by race. The effects of school accountability are largest among Hispanic students with 12 years of exposure to accountability leading to a 7.7 percentage point increase in college attendance on average. Black students are 3.5 percentage points more likely to graduate high school but 1.6 percentage points less likely to graduate from college with a Bachelor's degree under school accountability, although neither of these estimates are significantly different from the estimates for White students. Further, when examining heterogeneity by the strength of school accountability, I find that school accountability is most effective in conjunction with promotion gates in the form of an exit exam.

---

[32] These results include the states that Hanushek and Raymond (2005) reported as having report card accountability prior to NCLB that were not disputed by Dee and Jacob (2011).
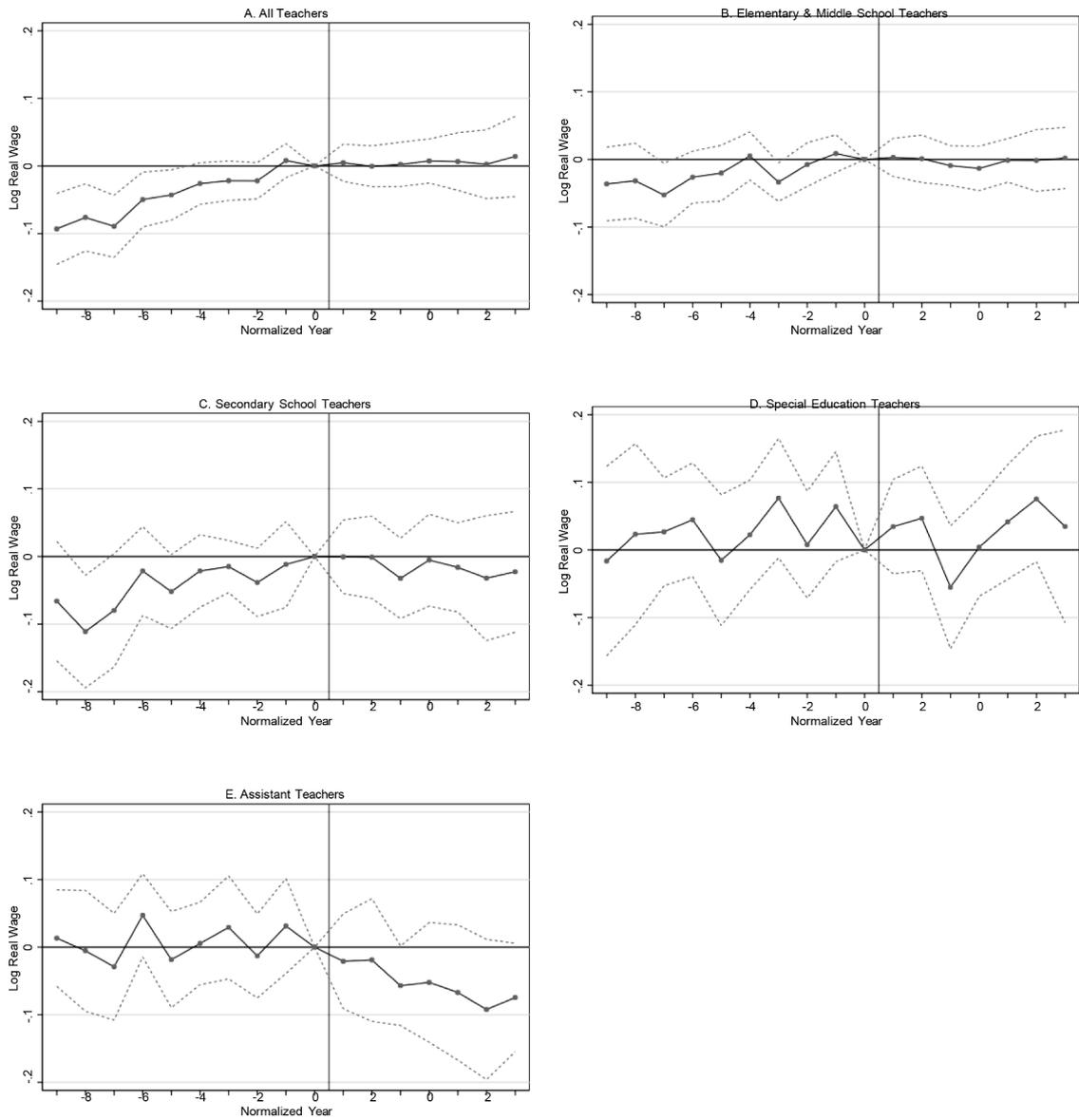
Fig. 9. Teacher wage.

Notes: Graphs of the log of real wage for teachers relative to the policy year. The regressions that produced this plot include dummy variables for years relative to the policy year and state, current age, and survey year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level. Common Core of Data School District Finance Survey (years 1990–2016).
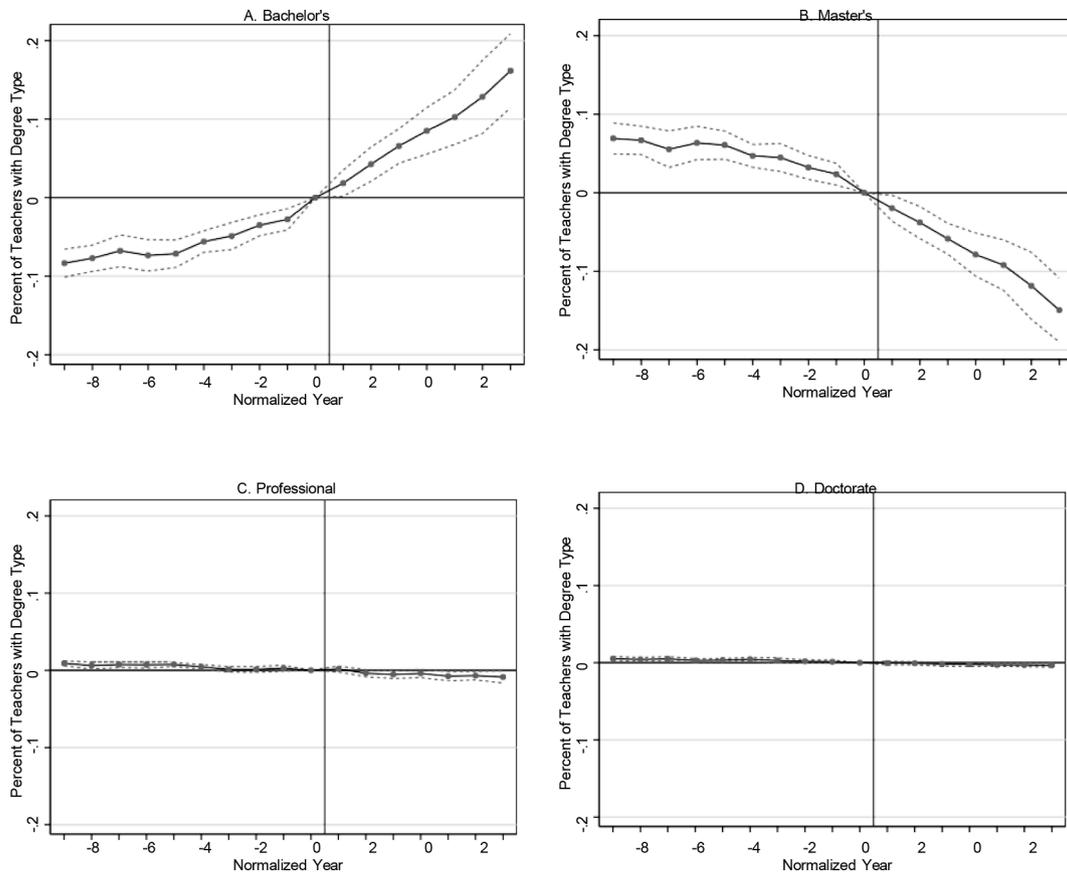
Fig. 10. Teacher degree.

Notes: Graphs for the degrees that teachers have relative to the policy year. The regressions that produced this plot include dummy variables for years relative to the policy year and state, current age, and survey year fixed effects. States that ever had report card accountability are excluded from this analysis. Standard errors are clustered at the state level.

Table 7
Balanced panel analysis.

Panel A: States (<= 1998 policy pass year)			
Dependent Variable	*High School*	*College Attend*	*BA Degree*
10 Years Exposure	0.003	0.006	0.004
	(0.005)	(0.006)	(0.005)
Full Exposure (12 years)	0.014*	0.009	0.014
	(0.007)	(0.007)	(0.012)
Adj. R-squared	0.024	0.042	0.057
N	2,573,410	1,901,829	1,597,100
Panel B: Observations (>= 8 y.o. in policy year)			
Dependent Variable	*High School*	*College Attend*	*BA Degree*
10 Years Exposure	0.025***	0.008	−0.003
	(0.008)	(0.006)	(0.006)
Full Exposure (12 years) 0.000	0.000 (.)	0.000 (.)	(.)
Adj. R-squared	0.025	0.043	0.055
N	4,118,526	3,463,139	3,093,760
Race, ethnicity, & gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes

Notes: Regression estimates for samples with a balanced panel using the fully flexible specification from Eq. (2). Panel A includes the same states for the whole analysis. Panel B keeps policy year ages with at least 300,000 observations. States that ever had report card accountability are excluded from this analysis. Robust standard errors are clustered at the birth state level.

* $p < 0.05$, ** $p < 0.01$,.
*** $p < 0.001$.

Table 8
Including expenditure reforms.

Dependent Variable	High School	College Attend	BA Degree
Full Exposure (12 years)	0.021**	0.009	0.012
	(0.008)	(0.007)	(0.011)
Race, ethnicity, & gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Adj. R-squared	0.025	0.043	0.055
N	4,828,380	3,736,937	3,215,494

Notes: Consistent estimates after controlling for school expenditure reforms. Full exposure is being exposed to school accountability for 12 years. States that ever had report card accountability are excluded from this analysis. Robust standard errors are clustered at the birth state level. * $p < 0.05$,.
** $p < 0.01$, *** $p < 0.001$.

Investigating the potential mechanisms behind the results found in this paper, I rule out state and district school expenditures, teacher wages, teacher qualifications, and the proportion of students placed in language assistance programs. I find suggestive evidence that schools are classifying a higher proportion of their students as eligible for special education. There are many other potential mechanisms through which school accountability can work including, but not limited to, curriculum, after-school programs, and time spent on core subjects. Investigating these other potential mechanisms is necessary to understand how schools are responding to school accountability and should be explored

Table 9
Including "Report Card" accountability states.

Dependent Variable	High School	College Attend	BA Degree
Full Exposure (12 years)	0.023***	0.013**	0.009
	(0.007)	(0.006)	(0.011)
Race, ethnicity, & gender controls	Yes	Yes	Yes
Birth-state fixed effects	Yes	Yes	Yes
Current age fixed effects	Yes	Yes	Yes
Survey year fixed effects	Yes	Yes	Yes
Adj. R-squared	0.025	0.043	0.053
N	5,297,835	4,163,358	3,586,381

Notes: Stable estimates after including states that had report card accountability prior to NCLB. Full exposure is being exposed to school accountability for 12 years. Report card accountability states are those that Dee and Jacob do not consider to have accountability until 2002 but that Hanushek and Raymond consider as having Report Card accountability prior to 2002. Robust standard errors are clustered at the birth state level. *$p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

in future research.

Since their adoption, and especially after NCLB, school accountability policies have been a topic of contention. Teachers and schools criticize the policies and feel the sanctions resulting from unsatisfactory student progress are unjust, while researchers and parents worry about the perverse incentives for teachers to teach to the test and to divide their attention unevenly across students. My findings indicate that, despite its controversy, school accountability has led to real improvements in educational attainment in the long run.

Funding

Research data

All data used in this article are publicly available.

CRediT authorship contribution statement

Katherine McElroy: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Declarations of Competing Interest

None.

Data availability

Data will be made available on request.

Acknowledgments

References

Ayers, J., Owen, I., Partee, G., & Chang, T. (2012). *No child left behind waiver applications: Promising ideas from second round applications.* Center for American Progress.

Baker, O., & Lang K., (2013). The effect of high school exit exams on graduation, employment, wages and incarceration (working paper 19182). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19182.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis, 24*(4), 305–331.

Cullen, J. B., & Reback, R. (2006). Tinkering toward accolades. *School gaming under a performance accountability system.* Emerald Group Publishing Limited.

Cullen, J. B., Jacob, B. A., & Levitt, S. D. (2005). The impact of school choice on student outcomes: An analysis of the Chicago public schools. *Journal of Public Economics, 89* (5–6), 729–760.

Dee, T. & Jacob B. (2006). Do high school exit exams influence educational attainment or labor market performance? (working paper 12199). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12199

Dee, T. S., & Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and management, 30*(3), 418–446.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics, 98*(5), 848–862.

Dinerstein, M., & Smith, T. (2021). Quantifying the supply response of private schools to public policies. *American Economic Review, 111*(10), 3376–3417.

Doan, K. P. (2008). No child left behind waivers: A lesson in federal flexibility or regulatory failure. *Administrative Law Review, 60*, 211.

Goertz, M. E., & Duffy, M. C. (2001). Assessment and accountability across the 50 states. *cpre policy briefs.*

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of policy analysis and management, 24*(2), 297–327.

Hanushek, E. A., & Rivkin, S. G. (2010). The quality and distribution of teachers under the no child left behind act. *Journal of Economic Perspectives, 24*(3), 133–150.

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics, 131*(1), 157–218.

Klein, A. (2015a). No child left behind: An overview. https://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definition-summary.html Last Access: 1/10/2020.

Klein, A. (2015b). No child left behind: An overview. *Education week, 34*(26), 1.

Krieg, J. M. (2008). Are students left behind? The distributional effects of the no child left behind act. *Education Finance and Policy, 3*(2), 250–281.

Lee, J., & Orfield, G. (2006). Tracking achievement gaps and assessing the impact of NCLB on the gaps.

Levesque, E. M.. The long-term impact of NCLB waivers on esea renewal. Re- trieved from. https://www.brookings.edu/blog/brown-center-chalkboard/2015/12/10/the-long-term-impact-of-nclb-waivers-on-esea-renewal/.

Loeb, S., & Page, M. (2000). Examining the link between wages and quality in the teacher workforce: The role of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics, 82*(3), 393–408.

McIntosh, S. (2012). *State high school exit exams: A policy in transition.* Center on Education Policy.

NCLB. (2002). Law 107-110, the no child left behind act of 2001. In *In 107th Congress.* Retrieved June.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics, 92*(2), 263–283.

Ruggles, S., Flood S., Goeken R., Schouweiler M., & Sobek, M. (2020). Ipums usa: Version12.0 [dataset]. Minneapolis, MN: IPUMS, 2022. 10.18128/D010.V12. 0.

Schellenberg, J., & Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review, 110*(5), 1502–1539.

Taylor, J., O'Day, J., & Le Floch, K.C. (2010). State and local implementation of the "no child left behind act". volume ix–accountability under "nclb": Final report. US Department of Education.