Growth-based school accountability rules and grade retention practices

Will Labadie

November 16, 2020

Abstract

In this paper, I explore the effect of public school accountability rules on school-level grade retention policy. In particular, I study the impact of accountability criteria that are based on year-to-year changes in student test scores. I develop a simple framework to simulate an administrator's retention strategy decision under a test score level-based accountability system and under a system that evaluates schools based on both levels and changes. The baseline framework predicts that in order to maximize the school-wide average test score, an administrator must retain almost all failing students; to maximize both the school-wide average level and the school-wide average increase in scores, an administrator must promote more aggressively in later grades. Students promoted from the last grade of a school are removed from the pool of test-takers, and the framework consistently predicts low retention in the final grade offered by a school after test score changes are added to the accountability system. Using a novel dataset of school-grade level retention rates for 7 states in the U.S. and an event study design, I find that about 30% fewer students are retained on average each year when a state adds student growth to the accountability criteria by which schools are evaluated. This number roughly corresponds to around 360,000 fewer retained students each year nationwide, and $4.2 billion saved in public school expenditures. I further find that administrators do retain significantly fewer students in the last grade offered by their schools. My results suggest that school administrators are willing to use retention as a tool for optimizing their schools' accountability ratings, and that the components of accountability systems do alter administrator behaviors.

1 Introduction

Each school year, public school administrators across the country endeavor to guide their schools to satisfactory ratings as defined by their state's adopted accountability system. What tools an administrator might use and how she might use them likely depend on what ratings are based on. To the extent that administrators decide whether or not to retain a student with an eye towards optimizing their schools' ratings, different ratings criteria could lead to different decisions.

This paper evaluates the impact of a common accountability measure, student growth, on public school retention policy. Retention has been shown to have a profound effect on students in the short and long term, but the determinants of retention are not well-understood (Schwerdt, West and Winters, 2017; Eren, Lovenheim and Mocan, 2018). This paper sheds light on how administrator incentives could affect the decision.

This paper shows that administrators do change their retention practices in the presence of growth-based accountability criteria. Almond, Lee and Schwartz (2016) show that in New York City, public school administrators did exercise influence over the retention decision, selectively enforcing a test score-based retention policy depending on students' race, poverty status, and previous performance. My paper shows that administrators retain students differently depending on the accountability criteria they face. A large literature exploring the unintended consequences of accountability systems has established already that administrators value satisfactory ratings (Rockoff and Turner, 2010). Research in this literature has shown that administrators use numerous tools to improve their schools' ratings, and that the interventions employed by schools can be targeted at the individual level (Figlio and Winicki, 2005; Reback, 2008). This paper establishes retention as one of those tools. Cullen and Reback (2006) find evidence of schools strategically exempting certain students from taking state exams to improve their ratings. Through retention, an administrator can alter the body of students that contribute to his school's rating - particularly in the last grade offered by the school.

This paper sheds light on the view that administrators take of retention's effect on student development. The literature investigating the effects of retention is broad and the results mixed. Taken together, the weight of the current evidence suggests that retention yields modest short-term gains to student achievement, particularly at young ages. Jacob and Lefgren (2004) find that retention substantially improved academic achievement for third graders and had little effect on sixth graders;Schwerdt, West and Winters (2017) found that in Florida, retention in third grade substantially improved short-term achievement. Nunes, Reis and Seabra (2018)find that among Portugese fourth graders, retention had a small positive effect on sixth grade achievement. Short-term gains are borne out in achievement data from Texas; in Texas public schools from 2012-2018, retained students nearly tripled their math exam passing rate after being held back. However, research also largely suggests that retention has substantial negative long-run effects on students. In the long term, retention has been shown to decrease students' likelihood of completing high school when applied in the eighth grade (Jacob and Lefgren, 2009); increase likelihood of student dropout and decrease subsequent achievement (Manacorda, 2012); to increase an individual's likelihood to be convicted of a crime by age 25 (Eren, Lovenheim and Mocan, 2018); and suppress an individual's entry wage (Brodaty et al., 2013). An administrator weighing the potential costs and benefits of retention likely considers both the long- and short-term consequences.

To simulate the retention decision, I develop a statistical framework in which students' test scores progress over time and administrators set score thresholds for each grade; students that score below the relevant score threshold are retained, while those that score above the threshold are promoted. This framework predicts different score threshold patterns for maximizing school-wide passing rate and school-wide growth, with the difference coming in the final grade offered by a school.

To estimate the impact of growth-based accountability criteria on retention, I exploit the adoption of student growth into seven states' accountability criteria over time. I use a difference-in-differences strategy and a novel dataset assembled with the help of the states' education agencies to evaluate the effect of these accountability system changes. While the timing of states adopting the growth criteria is not random, it is exogenous to any school-level decision-making. Following Goodman-Bacon (2018), I employ an event-study design in my analysis and include the balance test suggested by the author. I study the differential effect in the final grade offered by a school with a triple-differences technique.

I find that in the presence of growth-based accountability criteria, fewer public school students are held back, and that the effect grows over time. After 6 or more years under an accountability system that includes growth, retention rates decrease by 0.589 percentage points on average each year. This represents an 18.3% decrease relative to the average whole-sample retention rate. A back-of-the-envelope calculation using average enrollment counts from the Common Core of Data suggests that these estimates correspond to around 127,000 fewer retained students each year nationwide. The reduction in years in the school system results in an average of $1.8 billion less in student expenditures.[1]

I find that the effect of growth-based accountability criteria on retention is more pronounced in the final grade offered by a school. This result is consistent with the predictions of my statistical framework, and holds when the sample is restricted to elementary, middle, and high schools separately. This result suggests strategic retention on the part of school administrators, as the decision to retain or promote in the terminal grade affects the pool of test-taking students.That the differential effect of growth-based accountability criteria is negative suggests that administrators view retention as potentially harmful to year-over-year student score growth.

Finally, I find some evidence that male students are especially likely to be held back as a result of the adoption of growth-based accountability criteria based on data from Texas. Using a simple 2x2 difference-in-differences strategy, I find that the adoption of growth-based accountability criteria decreases the retention rate of males in terminal grades by .137 percentage points, and that the negative effect is consistent across school types. I find no evidence of differential effects by race, suggesting that administrators do not expect different effects by race on average.

[1] Based on current per-student expenditures of $13,847 (**?**).

The rest of the paper is structured as follows. Section 2 will elaborate on school accountability and the incentives of administrators. Section 3 lays out my statistical framework. Section 4 discusses the data I use to test the predictions of the framework. Sections 5 and 6 detail the empirical strategies I use and the results of the analyses respectively. Section 7 concludes.

2 Background

Since the enactment of the No Child Left Behind (NCLB) Act of 2001, states have been required to evaluate public schools based on student performance on standardized math and English Language Arts (ELA) exams. Under NCLB, state education agencies were required to evaluate schools based on the rate at which students passed its standardized exam, and on the rate at which various subpopulations of students at the school passed the exam. However, NCLB allowed state education agencies a substantial degree of flexibility in designing their accountability systems; each agency could choose its own standardized exam to administer, and define what level of performance would be considered adequate, and accountability systems could and did evaluate schools on criteria beyond the minimum competency requirement. A common criteria, especially among states that had accountability systems in place prior to the enactment of NCLB, was year-to-year growth in individual students' scores.

The Every Student Succeeds Act (ESSA) of 2015 requires state education agencies to develop accountability systems that include multiple evaluation criteria, which must include "students' performance on the statewide assessment, high school graduation rates, and English language proficiency" as well as at least one additional measure of school quality that is left up to the individual agencies (Alliance for Excellent Education, 2016).

Figure 1 shows the percent of states whose accountability systems included a measure of student test score growth from the 2004-05 to 2017-18 school years. In the 2004-05 school year, 9 states evaluated schools partially based on the growth rate of students; by the 2017-18 school year, every state did. While states did not randomly choose whether or not to adopt growth into their accountability systems, the decision was exogenous to school-level decision-making, and affected the way in which schools were evaluated.

Clearly, a school administrator's incentives are affected when the measures by which his school is evaluated change. Once student score growth is included in a school's rating, the administrator must attempt to satisfy the state-defined criterion along with the other indicators the state education agency considers in order to maximize his school's overall rating. A deep body of research has shown that authorities at schools use a number of tools to increase their school's overall rating. Reback (2008) finds that Texas students whose scores are relatively important to their school's overall rating perform better than expected, and finds evidence of finely targeted resource allocation and instruction to these "high-leverage" students; Figlio and Winicki (2005) find that schools in Virginia at risk of accountability-based sanctions increase the caloric content of school lunches on test days; Craig, Imberman and Perdue (2013) find that district administrators increase instructional budgets after facing a negative rating shock. While Cullen and Reback (2006) find that Texas schools manipulate the pool of test-takers via exemptions, no study has examined whether or not administrators hold students back strategically as well.

Under a minimum-competency based accountability system, a ratings-motivated administrator might hold a student back for a number of reasons. If the student is in grade g and the administrator expects the student to pass the grade g exam in the coming school year, but not the grade $g+1$ exam, he might retain the student to bolster the school's passing rate. This is particularly true if the student is from one of the subpopulations whose passing rates are explicitly valued under NCLB. The administrator might also retain the student if he believes that retention will positively impact the student's development, and have a lasting and positive impact on the student's future scores. Some research has shown that retention has positive short-run effects on student scores, but negative long-run effects on student development (Schwerdt, West and Winters, 2017); if a school is very likely to fail and its administrator estimates that the short-run effect of the student passing after being retained exceeds the later drag the student might have on the school's passing rate. On the other hand, if the administrator estimates that the long-run drag exceeds the short-run benefit, he would be more likely to promote the student.

Under an accountability system that includes year-to-year growth, the incentives to retain change. On one hand, an administrator might be more likely to retain a student if he expects the short-run benefits to be hight enough; it would improve his school's likelihood of passing both the state's minimum competency and growth standards. On the other hand, if retention has persistent negative effects, holding a student back will make it less likely that his school passes the state's growth standard. Because the overall direction is unclear, in section 3 I build a simulate a basic model of grade retention.

3 Statistical framework

In this section, I build a simple framework in which an administrator of a 3-grade school observes student scores each year and sets score-based promotion thresholds that determine which students will be promoted out of their current grade and which students will be retained in their current grade. I formulate a student score accumulation function of the following form:

$$S_{it} = \theta_i + \epsilon_{it} + \sum_{t=1}^{T} \gamma_{it}, \tag{1}$$

where S_{it} represents student i's test score in year t; $\theta_i \sim$ N(0,1). Students have some initial ability level θ_i that determines their place in the initial distribution of scores in a school, and their ability grows by γ_{it} each year.[2] ϵ_{it} represents a random shock to a student's score, which includes various unobserved components contributing to a student's score including the imperfect way in which ability maps to scores.[3] N students

[2]In the baseline case, $\gamma_{it} \sim$ N(0.5,0.25).

[3]In the baseline case, $\epsilon_{it} \sim$ N(0,0.5).

enter grade 1 of the school each year, and once students are promoted out of grade 3, they exit the school and their scores no longer contribute to the school's passing or growth rates. The school administrator sets promotion thresholds for each grade, δ_1, δ_2, and δ_3, that determine whether students are promoted or retained. The administrator seeks to maximize either the school's overall passing rate, $\frac{\sum_{g=1}^{3}\sum_{i=1}^{N_{gt}} \mathbb{1}(S_{it} \geq \pi_g)}{\sum_{g=1}^{3} N_{gt}}$, where N_{gt} represents the number of students in grade g in year t and π_g represents the score above which students in grade g must score in order to pass the standardized exam; or the administrator will seek to jointly maximize the school's passing rate and growth rate, $\frac{\sum_{g=1}^{3}\sum_{i=1}^{N_{gt-1}} \mathbb{1}((S_{it}-S_{it-1}) \geq \lambda_g)}{\sum_{g=1}^{3} N_{gt-1}}$, where N_{gt-1} represents the number of students in grade g of the school in year $t-1$ and λ_g represents the amount of score growth for students in grade g that is deemed adequate by the state.

3.1 Simulation results

A Monte Carlo exercise in which I perform 10,000 simulations yields some predictions about administrator behavior under the parameters assumed above. Figure 2 summarizes the results of the simulation, plotting the profile of δ_1, δ_2, and δ_3 cutoffs that maximize the school's passing and growth rates on average across the simulations. The results indicate that in the base case, an administrator attempting to maximize only passing rate should retain all failing students in grades 1 and 2, and should set the δ_3 threshold lower in grade 3, promoting students that passed and those that nearly passed. This strategy yields the highest average passing rate across simulations. An administrator interested in maximizing both passing and growth rate, on the other hand, should retain dramatically fewer students in grade 2 and especially in grade 3 relative to the passing-rate-only case.

4 Data

I use a novel source of administrative data I have collected from 8 states' education agencies. The data includes school-grade level retention rates and information on the components of each state's accountability ratings system from the 2011-12 school year to the 2017-18 school year. I also rely on the Common Core of Data (CCD) in conjunction with this data for enrollment counts for various subpopulations in the school in each grade, the grades offered by each school, Title I status, charter status, and magnet status. Overall, my data covers around 12,400 public schools across 7 years in which states gradually adopt growth in their accountability ratings systems.

For a state accountability system to evaluate schools based on growth can mean different things. Most states that include growth in school ratings measure the percent of students in each school whose test score increased by a sufficient (what "sufficient" means is defined by the state agency) amount from one year to the next; some states define growth as the difference between a student's score and a comparable (again, "comparable" is defined by the state agency) group of students' scores from the previous year, to attempt to average out idiosyncracies in test performance. However, all definitions emphasize year-to-year score increases.

5 Empirical analysis

5.1 Measuring the Impact of Growth-based Accountability Systems on Retention Practices

To study the effect of growth-based accountability systems on retention practices, I use an event-study design of the following form:

$$\begin{aligned} r_{gcdst} = &\alpha + \beta_1 \mathbb{1}(t - T_s^* \leq -2)_{st} + \beta_2 \mathbb{1}(t - T_s^* \in [0,2])_{st} \\ &+ \beta_3 \mathbb{1}(t - T_s^* \in [3,5])_{st} + \beta_4 \mathbb{1}(t - T_s^* \in [6,8])_{st} \\ &+ \gamma_c + \eta_g + \epsilon_{gcdst}. \end{aligned} \tag{2}$$

Here, r_{gcdst} represents the retention rate in grade g of campus c in district d of state s, year t, and T_s^*: year in which state s adopted student score change component in school ratings criteria. γ_c and η_g represent school and grade fixed effects respectively. In all specifications, I cluster my standard errors at the state-year level, since treatment is assigned at that level (Abadie et al., 2017).

The event study design exploits the fact that schools in different states are exogenously exposed to a growth-based evaluation system at different times. I rely on this design to identify the effect of treatment over time; the design also avoids bias that would be present in a simple difference-in-differences design if the treatment effect does change over time (Goodman-Bacon, 2018). The results of a placebo test show no pre-treatment effect and can be found in the first row of Table tk.

The administrator's decision is clearest in the last grade offered at his school. If he expects a student to perform poorly if held back, promoting the student removes him from the pool of students whose test scores determine the school's rating. To test the effect of growth-based accountability systems on retention practices in the last grade offered by a school, I estimate equation 2 separately for each grade, restricting attention to the first grade offered by a school, the middle grades offered, and the last grade offered. I also include a dummy variable identifying a grade as the last offered by a school as a third difference, estimating the following:

$$\begin{aligned} r_{gcdst} = &\alpha + \beta_1 \mathbb{1}(t - T_s^* \geq 0)_{st} \times \mathbb{1}(g = G_{ct}^T) + \beta_2 \mathbb{1}(t - T_s^* \geq 0)_{st} \\ &+ \beta_3 \mathbb{1}(g = G_{ct}^T) + \gamma_c + \epsilon_{gcdst}, \end{aligned} \tag{3}$$

where G_{ct}^T represents the final grade offered at school c in year t. The results of a placebo test show no statistically significant pre-treatment effect, and can be found in row 1 of Table tk.

5.2 Effect heterogeneity

If a school administrator has certain beliefs about how different subpopulations might perform after being retained, they may be differentially impacted by the switch to growth-based accountability. In addition, since the performances of some subpopulations are heavily weighted in many accountability systems, the policy impact on retention practices may differ by subpopulation. To test for differential effects, I utilize two different approaches.

First, I consider only the case of Texas. The Texas Education Agency makes retention rates available at the subpopulation-grade-year level; as a result, I am able to test for the effect of Texas' switch to a growth-based accountability system on subgroup-specific retention rates. To do so, I use a simple difference-in-differences design. Assignment to treatment depends on the year of observation - Texas school ratings included a measure of student growth starting in the 2012-13 school year - and I use an indicator for whether or not a given grade is the last offered by a school as a measure of treatment intensity. Thus, I am comparing the retention rates of various subgroups in the last grade offered by a school to those of the same groups in all other grades offered by the school before and after exposure to a growth-based accountability system. I estimate regressions of the following form:

$$\begin{aligned} r_{pgct} =& \alpha + \beta_1 \mathbb{1}(t \geq 2013)_t \times \mathbb{1}(g = g_c^T)_{gct} + \beta_2 \mathbb{1}(t \geq 2013)_t \\ &+ \beta_3 \mathbb{1}(g = g_c^T)_{gct} + n_{pct} + \eta_c + \epsilon_{gct} \end{aligned} \tag{4}$$

where r_{pgct} represents the retention rate of subgroup p students in grade g at campus c in year t, g_c^T represents the last grade offered at campus c and n_{pct} represents the total number of sudents of subgroup p enrolled in school c in year t. I estimate this equation separately for elementary and middle schools.[4] The validity of this approach requires that retention rates in terminal and non-terminal grades exhibiting parallel trends in the pre-treatment period. Most subgroups of interest fail an informal parallel trend check; for this reason, I focus only on male, Black, Hispanic, and white students in this analysis. Figure 4 includes plots of the retention rate for each of these subgroups over time, from 2004-05 to 2017-18, in terminal and non-terminal grades.

The obvious drawback of the Texas-only approach is that it identifies a treatment effect in only one state. The treatment effects identified by that analysis are Texas-specific, and are a result of all the components and rule changes that came with the 2011-12 accountability system change. To assess the various potential impacts of growth-based accountability systems on the retention of different subpopulations, I use the Common Core of Data (CCD) and a blunt measure of retention. By using the CCD, I am able to expand the scope of analysis; I use data from all 50 states and Washington, D.C. However, I must construct a proxy for retention, as it is not included in the CCD. If an administrator retains a certain subgroup more after the switch to growth-based accountability, the number of students from that subgroup enrolled in a given grade should increase relative to the grade's pre-treatment average. Similarly, if the administrator is more likely to promote a student of a given subgroup after the switch to growth-based accountability, the enrollment count of that subgroup should fall relative to the pre-treatment average. Based on this logic, I use "excess enrollment" - the difference between the number of students of a given subgroup enrolled in a given

[4]High schoolers that repeat a course are counted as retained by TEA; for this reason, I choose to exclude them from my analysis.

grade and the pre-treatment average number of students of that subgroup enrolled in that grade - to measure retention. I estimate regressions of the following form:

$$\begin{aligned} n_{pgcdst} - \bar{n}^{\text{pre}}_{pgcdst} =& \alpha + \beta_1 \mathbb{1}(t - T^*_s \leq -2)_{st} + \beta_2 \mathbb{1}(t - T^*_s \in [0,2])_{st} \\ &+ \beta_3 \mathbb{1}(t - T^*_s \in [3,5])_{st} + \beta_4 \mathbb{1}(t - T^*_s \in [6,8])_{st} \\ &+ \gamma_c + \delta_{dt} + \epsilon_{pgcdst} \end{aligned} \quad (5)$$

Here, $n_{pgcdst} - \bar{n}^{\text{pre}}_{pgcdst}$ represents the "excess enrollment" of subgroup p in grade g of school c in district d of state s in year t, and T^*_s represents the year in which state s incorporated student score growth into school ratings. To control for local population trends, I include district-year fixed effects.

6 Results

6.1 The effects of Growth-based Accountability on retention rates

Table 1 presents the overall estimated effects of growth-based accountability on retention rates. Column (1) presents the results of estimating equation 2 without grade fixed effects, and column (2) includes them; column (2) represents my preferred specification. The results suggest modest negative effects on retention rates overall in the first few years after the adoption of growth-based accountability measures, though the estimated effects in these years are not statistically significant. After perhaps an adjustment period, retention rates fall by 0.620 percentage points - a 30% decrease from the pre-growth average.

Table 2 presents the main results on the effect of growth-based accountability on retention rates in the final grade offered by schools. The results suggest negative effects of a similar magnitude to the overall effect found in Table 1 in each post-treatment period; perhaps because the impact of promoting a student out of school is certain, the impact of the policy was more immediate in the terminal grades offered by schools.

6.2 Effect heterogeneity

Tables 3, 4, 5, and 6 present the effects of growth-based accountability on the retention rates of male, Black, Hispanic, and white students respectively in Texas. I estimate equation 4 separately for elementary, middle, and high schools, as retention rates vary considerably across the three school types. The results presented in Table 3 show that the switch to growth-based accountability had a relatively small but significant effect on the retention rate of elementary school boys. The retention rate among boys in terminal grades of decreases by .137 percentage points after the switch to growth-based accountability - a 4% decrease relative to the pre-growth average retention rate. This is largely fueled by a decrease in the retention rate of high school boys in terminal grades (i.e. high school graduates), and partly by a decrease in the retention rate of elementary school boys. The rate at which elementary school boys are held back decreases by .056 percentage points after the switch to growth-based accountability - a 1.6% decrease relative to the pre-growth average retention rate. The results of Tables 4, 5, and

6 show similar patterns - modest decreases in terminal-grade retention rates after the shift to growth-based accountability. Notably, the decreases for Black and Hispanic students seem to be driven primarily by decreases in the terminal-grade retention rate in middle school rather than in elementary or high school. Overall, I do not find compelling evidence that any of these subgroups are more affected by the policy change than others.

7 Conclusion

This paper provides evidence that public school administrators use retention as a tool in optimizing their school's rating. They respond to changes in the way the school is evaluated - particularly a switch to school ratings that incorporate student growth - by changing they rate at which they retain students. The point estimates based on a 7-state sample suggest that schools may retain 30% fewer students after operating under a growth-based system for at least 6 years, and that the effect may be more pronounced in the final grade offered by a school, where promotion of a student evicts them from the school's pool of test-takers. I find some suggestive evidence that Black and Hispanic students are more likely to be promoted out of the last grade offered by their middle school after a switch to growth-based accountability, and that boys are more likely to be promoted out of the last grade offered by their elementary school after the switch. These results suggest that administrator expectations about the effects of retention may differ by gender, age, and race. Overall, administrators choose to retain less when student score growth matters to their school ratings, particularly when they are able to remove the student from the test-taking pool via promotion, suggesting that administrators expect retention to be harmful to a student's score growth in the short term.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?"

Alliance for Excellent Education. 2016. "Every Student Succeeds Act Primer: Personalized Learning." Alliance for Excellent Education April, Washington, D.C.

Almond, Douglas, Ajin Lee, and Amy Ellen Schwartz. 2016. "Retention Heterogeneity in New York City Schools." , (November): 1–25.

Brodaty, Thomas O, Robert J Gary-Bobo, Ana Prieto, We Thank, Denis Fougere, Francis Kramarz, Guy Laroque, Edwin Leuven, Bentley Macleod, Thierry Magnac, Eric Maurin, and Jean-Marc Robin. 2013. "Does Speed Signal Ability? The Impact of Grade Retention on Wages." , (February): 1–54.

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.

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System." *Advances in Applied Microeconomics*, 14(May): 1–34.

Eren, Ozkan, Michael F. Lovenheim, and Naci H. Mocan. 2018. "The effect of grade retention on adult crime: Evidence from a test-based promotion policy."

Figlio, David N., and Joshua Winicki. 2005. "Food for thought: The effects of school accountability plans on school nutrition." *Journal of Public Economics*, 89(2-3): 381–394.

Goodman-Bacon, Andrew. 2018. "Difference-in-Differences With Variation in Treatment Timing." *National Bureau of Economic Research*, 17(5): 684–694.

Jacob, Brian A., and Lars Lefgren. 2004. "Remedial education and student achievement: A regression-discontinuity analysis." *Review of Economics and Statistics*, 86(1): 226–244.

Jacob, Brian A., and Lars Lefgren. 2009. "The Effect of Grade Retention on High School Completion." *American Economic Journal: Applied Economics*, 1(3): 33–58.

Manacorda, Marco. 2012. "The cost of grade retention." *Review of Economics and Statistics*, 94(2): 596–606.

Nunes, Luis Catela, Ana Balcão Reis, and Carmo Seabra. 2018. "Is retention beneficial to low-achieving students ? Evidence from Portugal." *Applied Economics*, 50(40): 4306–4317.

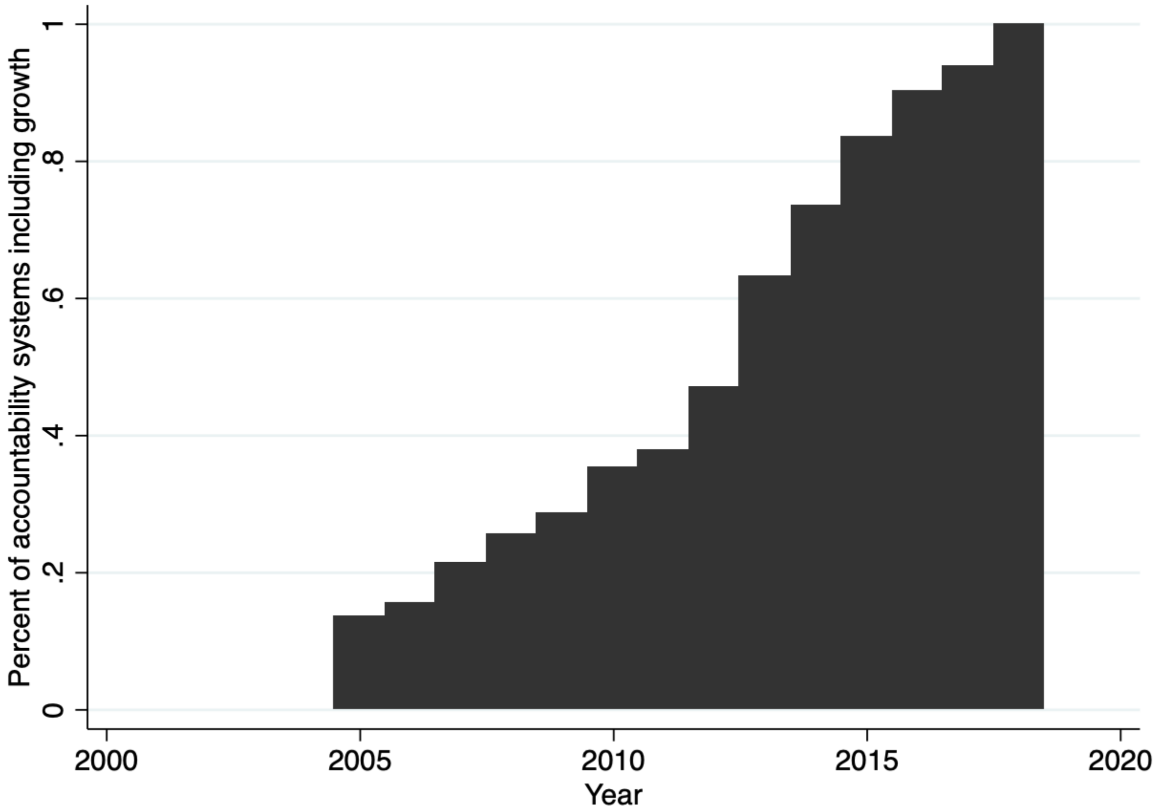
Reback, Randall. 2008. "Teaching to the rating : School accountability and the distribution of student achievement." 92: 1394–1415.

Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-run impacts of accountability on school quality." *American Economic Journal: Economic Policy*, 2(4): 119–147.

Schwerdt, Guido, Martin R. West, and Marcus A. Winters. 2017. "The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida." *Journal of Public Economics*, 152: 154–169.

8 Tables & Figures

Figure 1: Presence of growth in accountability systems over time



Notes: Includes all 50 states and Washington, D.C.
Sources: Education Week Education Counts Research Center; individual states' education agencies.

Figure 2: Passing and growth rate-maximizing δ profiles

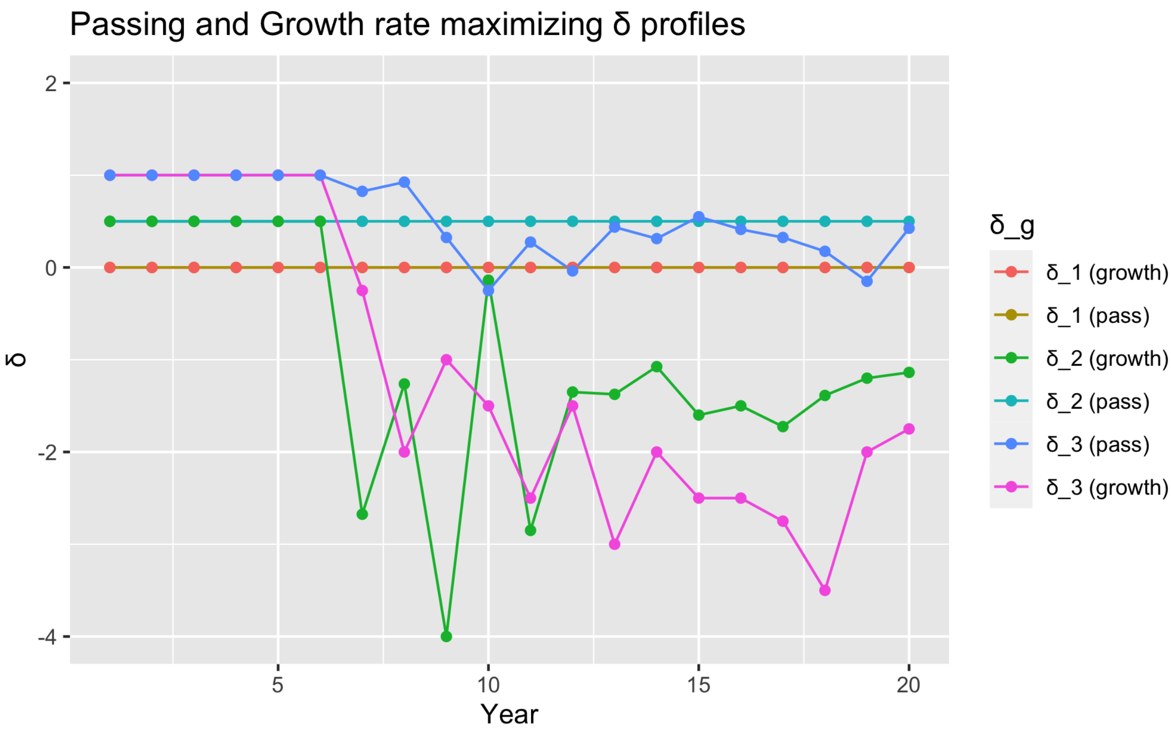
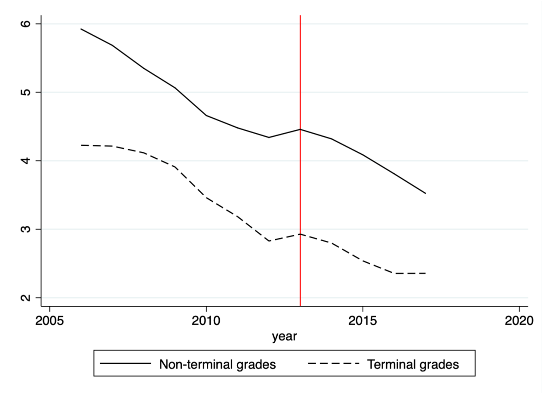
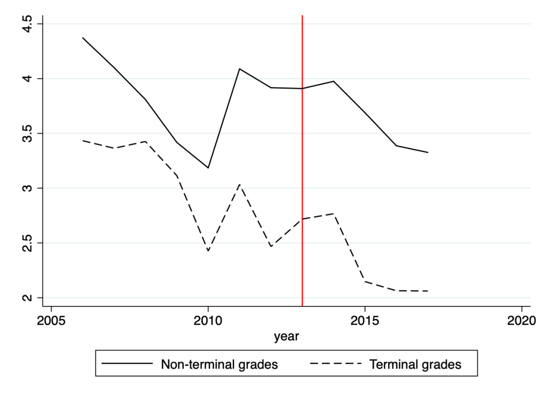


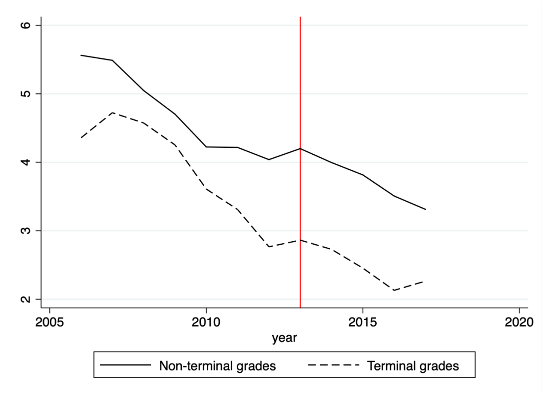
Figure 3: Parallel trend checks



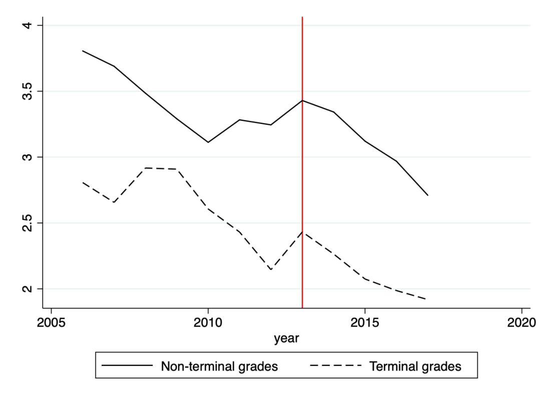
(a) Retention rate among male students



(b) Retention rate among Black students



(c) Retention rate among Hispanic students



(d) Retention rate among white students

Table 1: Event study results

	Dep. variable is grade-level retention rate	
	(1)	(2)
Years -5 to -2	0.097	0.136
	(0.061)	(0.096)
Years 0 to 2	-0.060	-0.087
	(0.080)	(0.085)
Years 3 to 5	-0.092	-0.122
	(0.087)	(0.092)
Years 6 to 8	-0.589***	-0.620***
	(0.094)	(0.098)
N	289820	289820
Grade FE		✓

Notes: Standard errors, clustered at the state-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects.

Table 2: Terminal grade triple-difference

	Dep. variable is grade-level retention rate	
	(1)	(2)
Years -5 to -2	0.097	0.033
	(0.061)	(0.063)
Years 0 to 2	-0.060	0.053
	(0.080)	(0.079)
Years 3 to 5	-0.092	0.002
	(0.087)	(0.087)
Years 6 to 8	-0.589***	-0.517***
	(0.094)	(0.103)
$\mathbb{1}(g = G_c^T)$		-0.625***
		(0.008)
Years -5 to -2 $\times$ $\mathbb{1}(g = G_c^T)$		0.369***
		(0.011)
Years 0 to 2 $\times$ $\mathbb{1}(g = G_c^T)$		-0.687***
		(0.059)
Years 3 to 5 $\times$ $\mathbb{1}(g = G_c^T)$		-0.614***
		(0.041)
Years 6 to 8 $\times$ $\mathbb{1}(g = G_c^T)$		-0.524***
		(0.089)
N	289820	289820

Notes: Standard errors, clustered at the state-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects.

Table 3: Effect of growth-based accountability on retention rate among male students: Texas

	Dep. variable is grade-level retention rate			
	Full sample	Elementary school	Middle school	High school
Post-growth	-0.014	-0.009	-0.003	-0.130
	(0.016)	(0.009)	(0.045)	(0.121)
$\mathbb{1}(g = G_c^T)$	0.065**	0.102***	0.075*	-0.597***
	(0.024)	(0.019)	(0.031)	(0.142)
Post-growth $\times$ $\mathbb{1}(g = G_c^T)$	-0.137***	-0.056*	-0.075	-0.660*
	(0.037)	(0.023)	(0.051)	(0.286)
N	245702	181322	41247	23133

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects and control for the total enrolled male students at the school.

Table 4: Effect of growth-based accountability on retention rate among Black students: Texas

	Dep. variable is grade-level retention rate			
	Full sample	Elementary school	Middle school	High school
Post-growth	0.002	0.002	0.004	-0.001
	(0.002)	(0.001)	(0.003)	(0.011)
$\mathbb{1}(g = G_c^T)$	0.025***	0.008*	0.018	0.039
	(0.007)	(0.004)	(0.011)	(0.023)
Post-growth $\times$ $\mathbb{1}(g = G_c^T)$	-0.020**	-0.005	-0.011	-0.040
	(0.007)	(0.004)	(0.010)	(0.026)
N	309871	232045	52542	25284

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects and control for the total enrolled Black students at the school.

Table 5: Effect of growth-based accountability on retention rate among Hispanic students: Texas

	Dep. variable is grade-level retention rate			
	Full sample	Elementary school	Middle school	High school
Post-growth	-0.005	0.009*	0.018	-0.056
	(0.006)	(0.005)	(0.021)	(0.043)
$\mathbb{1}(g = G_c^T)$	0.025*	0.035**	0.056**	0.082
	(0.011)	(0.013)	(0.018)	(0.055)
Post-growth $\times$ $\mathbb{1}(g = G_c^T)$	-0.028*	-0.020	-0.060*	-0.056
	(0.012)	(0.011)	(0.027)	(0.074)
N	261830	197226	45884	18720

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects and control for the total enrolled Hispanic students at the school.

Table 6: Effect of growth-based accountability on retention rate among white students: Texas

	Dep. variable is grade-level retention rate			
	Full sample	Elementary school	Middle school	High school
Post-growth	0.003	-0.001	0.006	0.018
	(0.001)	(0.001)	(0.004)	(0.010)
$\mathbb{1}(g = G_c^T)$	0.021***	0.015*	0.012	0.083*
	(0.006)	(0.007)	(0.008)	(0.034)
Post-growth $\times$ $\mathbb{1}(g = G_c^T)$	-0.016**	-0.010	-0.004	-0.094**
	(0.006)	(0.006)	(0.010)	(0.034)
N	309790	237318	51782	20690

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. * denotes significance at 10%, ** denotes 5%, and *** denotes 1%. All regressions include school fixed effects and control for the total enrolled white students at the school.