Which Students Are Held Back When School Accountability Rules Change?

(Latest version can be found here)

William Labadie

March 9, 2021

PRELIMINARY - DO NOT CITE

1 Introduction

Being retained in grade has profound, long-term effects on students. A vast literature has investigated the effects of retention, with the most consistent finding being that the effects are heterogeneous across students (Fruehwirth, Navarro and Takahashi, 2016; Gary-Bobo, Goussé and Robin, 2016). However, relatively little is known about why school administrators advocate for certain students to be retained. This study investigates how administrators' retention strategy changes when the accountability rules their schools are subject to change.

Labadie (2021) finds that when a state adopts a growth-based accountability system - one which includes a measure of year-to-year growth in within-student standardized test scores - administrators respond by retaining 18% fewer students on average, and that the effect is even more pronounced in the final grade offered by a school (the terminal grade). Given the developmental effects of retention, understanding which students are differentially promoted or retained due to the policy change is crucial for any insight into its welfare effects. Does a change to growth-based accountability reduce the number of students retained near the passing threshold - those who are unlikely to benefit from retention relative to those near the bottom of the score distribution (Gary-Bobo, Goussé and Robin, 2016)?

This study would contribute to a large literature analyzing the unintended consequences of school accountability systems. Many studies have found that accountability systems often lead to administrators and teachers focusing resources on "high-leverage" students (Reback, 2008) or schools (Craig, Imberman and Perdue, 2013). High-leverage students are those who are relatively likely to pass the standardized exam with additional resources, and high-leverage schools are those that are relatively likely to earn a passing rating from the state education agency with additional resources. The distributional effects of these effects is of great interest; such incentives lead to both high- and low-performing students and schools receiving fewer resources relative to the no-accountability case. Growth-based accountability systems, by placing value on within-student progress, gained popularity as a solution to these unintended consequences. This study would offer one evaluation of their effectiveness in doing so.

I extend the conceptual framework developed in Labadie (2021) to allow administrators more flexibility in the retention decision. The overall predictions of the model are the same: adding growth-based accountability results in more retention in early grades, and less in the last grade offered by a school. However, the distributional predictions of the model suggest that in the presence of growth-based accountability, administrators will retain higher-performing students in early grades than under passing rate-based accountability, and only the lowest-performing students in the terminal grade. This prediction, if consistent with school behavior, would suggest that growth-based accountability is successful in redirecting administrator focus from students close to passing the exam to low-achieving students. Under passing rate-based accountability, the model predicts much more retention of such high-leverage students, particularly in the terminal grade.

I analyze the response of administrators to growth-based accountability using student-level data from Texas public schools. This data covers the universe of Texas public school students from the 1994-95 school year through the 2017-18 school year. I plan to use a triple-differences technique to analyze the effect of the policy change on students with differing test scores, in terminal grades relative to those not in terminal grades. Using individual-level data is key, as it allows me to observe the test scores and retention status of individual students; aggregate data, which does not include average score, or even passing rate, by grade and retention status, does not allow for any distributional analysis.

The paper is structured as follows. Section 2 conducts a brief literature review. Section 3 describes my conceptual framework and simulation results. Section 4 describes the data I intend to use in more detail. Section 5 shows some descriptive statistics and preliminary results. Section 6 describes my proposed empirical strategy, and Section 7 concludes.

2 Literature review

Previous research on retention is largely limited to inquiries regarding its long- and short-term effects on students. Studies of the unintended effects of school accountability systems are numerous, but frequently constrained in scope by data limitations. Much of the research on retention has focused on the long-and short-term effects on being retained on an individual. Many have found that being held back has negative long-term effects on a student, including on beginning-of-career wage (Brodaty, Gary-Bobo and Prieto, 2013); probability of dropping out and delaying graduation (Cockx, Picchio and Baert, 2017; Manacorda, 2012); and probability of criminal conviction (Eren, Lovenheim and Mocan, 2018). Some, including Diris (2016) and Manacorda (2012), have even found negative short-term academic effects. On the other hand, some studies have found evidence that retention can be beneficial, particularly in the short term. The mechanism through which Duflo, et al. (2011) find that tracking can be beneficial to students can be applied to retention. A student performing below his class median might benefit from retention because his ability level will be closer to the median than if he had been promoted; further, his erstwhile class may benefit from the teacher being better able to target her instructional level. Greene and Winters (2007) found that among failing Floridian students, students that were retained outperformed those that were promoted in the years following retention. Figlio and Özek (2019) find substantial benefits of early retention for English language learners. Schwerdt, West and Winters (2017) find long-lasting positive effects on test scores and GPA with no effect on probability of graduation or dropout. A large strand of literature finds heterogeneous effects. In particular, the evidence suggests that early retention can improve academic performance, while later retention can be harmful to performance and dropout probability (Jacob and Lefgren, 2004, 2009); that early retention harms students on average, but might benefit students that would receive additional parental and school investment (Fruehwirth, Navarro and Takahashi, 2016); and that retention in junior high school is beneficial to the academic performance of weakest students, but reduces the probability of entry into high school (Gary-Bobo, Goussé and Robin, 2016).

One study closely related to the proposed research is that of (Almond, Lee and Schwartz, 2016) in which the authors test for differential enforcement of retention policy in New York City schools. This work, however, does not explore the incentive structure that may lead administrators to enforce retention policies more strictly for a given group than for another; looks only at schools in New York City, which has a unique institutional background distinct from that of Texas; and leaves the question of administrator preferences un-addressed, concluding only that administrators employ "patterned discretion" in the usage of standardized test scores. My proposed research would complement the work of Almond, Lee and Schwartz (2016) by accounting for important institutional factors and developing theoretical explanations for the observed trends to shed light on the economic decision-making processes administrators employ when faced with a retention candidate.

This study fits into the literature on the economic decision-making of school administrators, and in particular extends the literature on the way that administrators respond to the incentives created by accountability systems. (Craig, Imberman and Perdue, 2013) find that Texas school districts make short-term investments in instruction after failing to meet the state accountability standard. Figlio and Winicki (2005) document that Virginia schools facing accountability sanctions increase the caloric content of lunch menus on test days. Reback (2008) shows that Texas students perform better when their scores are especially important to their schools' accountability ratings, suggesting targeted resource management. On the whole, the research clearly shows that schools and districts are very sensitive to whether or not they pass accountability standard, and that they will change their behavior in targeted and subtle ways to increase their chances of passing.

The established research shows that administrators are motivated by school accountability standards, and that they use a variety of tools to satisfy the standards. It is also well-established that retention, particularly at early ages, affects student performance and thus school performance, though the direction of the effect is still a matter of debate. Further, the evidence suggests substantial negative long-term effects for students. It is important to study the determinants of retention, and the role that accountability standards play in the retention decision. Administrators may target retention, focusing on the students whose scores are most likely to benefit their schools' accountability score after retention (for example, younger students that nearly pass the STAAR exam), or they may be less likely to retain any given student given the high cost and uncertain test performance returns of retention.

3 Conceptual framework

Because the overall effect of retaining a student is unclear, I build and simulate a basic one-school model of grade retention. The model relies on a factor model formulation of student skill accumulation and skill measurement by standardized exams, following the basic structure of Cunha and Heckman (2008). In each period t, students at the school earn a score on the state's standardized exam. Assume that there is only one subject exam, and only one score counts towards the school's rating. Assume that student scores are represented by the following dynamic factor structure:

$$s_{it} = \mu + \alpha(g_{it})\theta_{it} + \epsilon_{it} \tag{1}$$

where i represents the individual student, and θ_{it} is a dynamic factor for each student. $\alpha(g_{it})$ is the factor loading parameter, which translates latent ability θ_{it} into a measurable score, s_{it}. It is a function of the student's grade, g_{it}, reflecting the fact that students in different grades take tests of different difficulty levels; $\frac{\partial\alpha(g_{it})}{\partial g_{it}} < 0$, such that students with the same latent ability values will earn lower scores in expectation if they are placed in higher grades. ϵ_{it} represents a random shock to the measured scores, representing the noisiness of test scores in measuring ability.

I assume the following form for students' skills production:

$$\theta_{it} = \gamma_0 + \gamma_1\theta_{it-1} + \eta_{it} \tag{2}$$

where η_{it} is assumed independent across students and over time for the same students. I assume that students draw some initial value θ_{i0} prior to entering school. For simplicity, I assume that retention does not affect the skills production function. Research suggests that it would be more realistic to include retention in the skills production function, and this is a consideration for future work.

The administrator's objective is to allocate students across grades to maximize either the percent of students passing the state exam (passing rate) or both the passing rate and the percent of students exhibiting sufficient test score growth from one year to the next (sufficient growth rate). The school's passing rate in year t is given by

$$\Pi_t = \frac{\sum_{j=1}^{G} \sum_{i=1}^{N} \mathbb{1}(g_{it} = j) \times \mathbb{1}(s_{it} \geq \pi)}{N_t} \tag{3}$$

where G represents the final grade offered by the school and N represents the total number of students. π represents the externally set passing exam score. N_t represents the number of students whose scores count towards the school's rating in year t:

$$N_t = \sum_{j=1}^{G} \sum_{i=1}^{N} \mathbb{1}(g_{it} = j). \tag{4}$$

The school's sufficient growth rate is assumed to be measured as follows:

$$\Lambda_t = \frac{\sum_{j=1}^{G} \sum_{i=1}^{N} \mathbb{1}(g_{it} = j) \times \mathbb{1}(s_{it} - s_{it-1} \geq \lambda)}{N_t} \tag{5}$$

where λ represents the externally set target amount of score growth for a given student, and all other objects are as defined above. In reality, accountability measures of student growth vary across states; however, the percent of students demonstrating a sufficient level of score growth is a central component of all the growth-based accountability measures I analyze in this paper.

The school administrator is able to choose the difficulty of the exam that a student is exposed to in a given year through retention. The administrator can choose to retain a grade g student, and she will take the grade g exam rather than the more difficult grade $g+1$ exam. I assume that administrators make the retention decision by setting promotion thresholds δ_g for each grade. Grade g students that score at or above δ_g are promoted to grade $g+1$, while grade g students that score below δ_g are retained and repeat grade g. That is, administrators control students' grade level g based on the following:

$$g_{it} = g_{it-1} + \mathbb{1}(s_{it-1} \geq \delta_g). \tag{6}$$

I assume that retention is the only tool administrators have to influence their school's scores in this model; they can only affect Π_t and Λ_t through their δ_g choices. δ_g affects student scores, s_{it}, directly through $\alpha(g_{it}(\delta_g))$, but it also affects the population of grades g and $g+1$ through N_t, since g_{it} is a function of δ_g.

I assume that the school administrator has two important constraints. First, he must promote students that pass the standardized exam:

$$\delta_g \leq \pi_g \tag{7}$$

and he cannot retain a student in the same grade two years in a row

$$g_{it} \neq g_{it-2} \forall i. \tag{8}$$

These constraints are based on common rules adopted by school districts and state education agencies. The first has the additional feature of removing the option of retaining all passing students to maximize the school passing rate, which is appealing for the simulation of this model. I assume that these constraints hold for administrators regardless of the accountability system their schools are subject to.

An administrator of a school under a passing rate-based accountability system has the following optimization problem:

$$\max_{\delta_1,\delta_2,\delta_3} \Pi(\delta_1,\delta_2,\delta_3) \quad s.t. \quad \delta_g \leq \pi_g, \quad g_{it} \neq g_{it-2} \forall i. \tag{9}$$

while an administrator of a school under a passing rate and growth-based accountability system has the following optimization problem:

$$\max_{\delta_1,\delta_2,\delta_3} \Pi(\delta_1,\delta_2,\delta_3) + \Lambda(\delta_1,\delta_2,\delta_3) \quad s.t. \quad \delta_g \leq \pi_g, \quad g_{it} \neq g_{it-2} \forall i. \tag{10}$$

The mechanisms driving the administrator's decision are shown graphically in Figure 2. Figure 2a plots the relationship between a student's grade level, his latent ability, and his test score. Because the α function is decreasing in student grade level, the same level of latent ability will translate into lower scores for students in higher grades in expectation. A given student i with latent ability θ_{it} in year t earns a test score equal to $s_{it}(\theta_{it}, \alpha(g_{it} = j))$ if he is in grade j in year t. Figure 2a shows a student whose latent ability grows from θ_{it} in year t to θ_{it+1} in year $t+1$, and the scores he would earn in year $t+1$ if he is retained or promoted. $\bar{s}$ represents the maximum achievable score. If retained, the student will earn a score of $s_{it+1}(\theta_{it+1}, \alpha(g_{it+1} = j))$; if promoted, his score will be substantially lower, at $s_{it+1}(\theta_{it+1}, \alpha(g_{it+1} = j+1))$. If π, the externally-set passing standard, is between $s_{it}(\theta_{it}, \alpha(g_{it} = j))$ and $s_{it+1}(\theta_{it+1}, \alpha(g_{it+1} = j+1))$, the administrator's objective function is expected to be as well-served by promoting the student as by retaining him. In this case, an optimal δ_1 will be set below $s_{it}(\theta_{it}, \alpha(g_{it} = j))$. If π is between $s_{it+1}(\theta_{it+1}, \alpha(g_{it+1} = j+1))$ and $s_{it+1}(\theta_{it+1}, \alpha(g_{it+1} = j))$, however, the administrator would be best-served by retaining the student. If promoted to grade $j+1$, the student would not be expected to pass the grade $j+1$ exam; if retained, the student would be expected to pass the grade j exam.

Figure 2b shows the relationship between a student's grade level, his latent ability, and the rate of change in his scores. For the same values θ_{it} and θ_{it+1} plotted in Figure 2a, Figure 2b plots the partial derivative of s with respect to θ. The shape is determined by the functional form of α, since α is a function of grade level, which is determined by score - a function of θ. If λ, the externally-set sufficient growth standard, is below $\frac{\partial s_{it}(\theta_{it}, \alpha(g_{it}=j+1))}{\partial \theta}$, then the student's score is expected to grow enough to pass the standard if he is promoted from grade 1 to grade 2. If λ is above $\frac{\partial s_{it}(\theta_{it}, \alpha(g_{it}=j+1))}{\partial \theta}$, the student's score growth is expected to satisfy the standard if he is retained, but not if he is promoted to grade $j+1$. To maximize the sufficient score growth standard only, administrators should set δ_j such that students with $\frac{\partial s_{it}(\theta_{it}, \alpha(g_{it}=j+1))}{\partial \theta} < \lambda$ are retained;

this occurs for students with relatively low θ_{it}. As a result, students with relatively low scores are those more likely to satisfy the sufficient growth standard when retained.

To gain some insight into the retention decision under different accountability systems, I perform a Monte Carlo exercise, in which I perform 10,000 simulations of the model. Each iteration simulates one school with three grades for ten years, with 100 new students entering the first grade each year. The parameter assumptions I employ are given in Table 1. I define the factor loading $\alpha_{it} = \frac{1}{\sqrt{g_{it}}}$, which is the functional form used in Figure 2. I assume that measurement is relatively noisy while the noise in skill accumulation is relatively small: $\epsilon_{it} \sim$ N(0,20). and $\eta_{it} \sim$ N(0,10). I also assume $\theta_{i0} \sim$ N(50,25), and that raw ability grows by 35 points per year in expectation, and thus assign $\gamma_0 = 35$. This means that the average student entering the school will score a 50 on the first grade exam in expectation, and if promoted each year, she will score a 60 on the second grade exam and a 69 on the third grade exam. Finally, I assume that the state passing standard, π, is equal to 50, such that the average student would be expected to pass the exam each year without needing to be retained. I assume that the state sufficient growth standard, λ, is equal to 10, about the rate at which the average student's scores are expected to naturally increase in the absence of retention. An administrator of a school with students scores centered substantially below or above the passing standard, or one whose student population's scores are expected to grow substantially slower or faster than the sufficient growth standard, could behave differently than the one I simulate.

Figure 1: Level and year-to-year change in student score are affected by grade level

Figure 2a: Student scores for two different grades

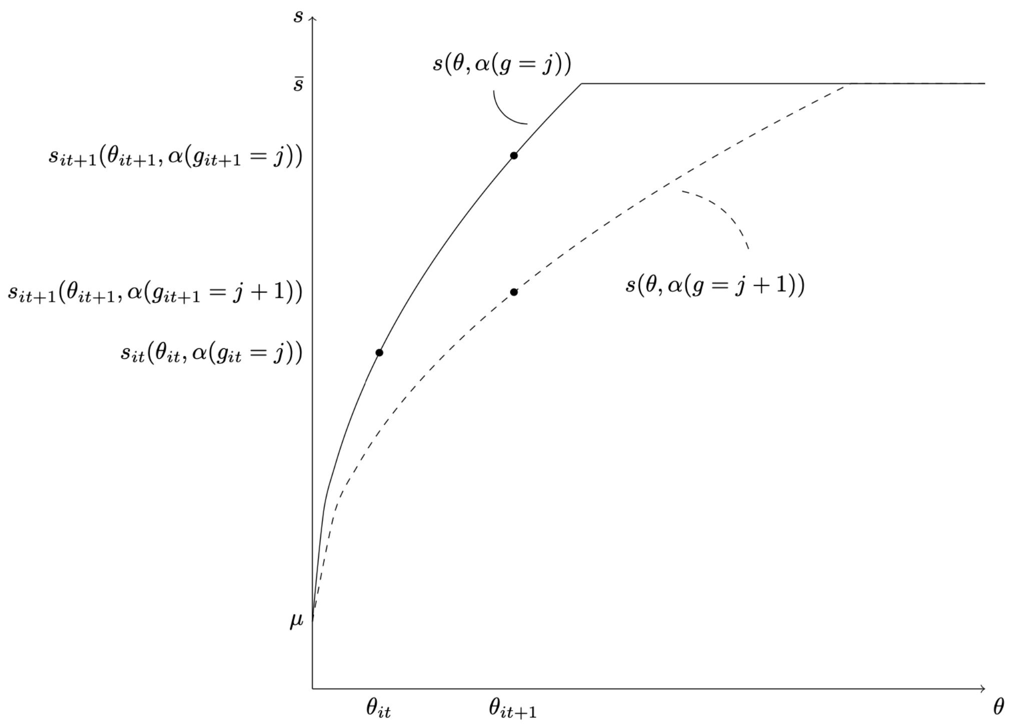


Figure 2b: Rate of change in student scores for two different grades

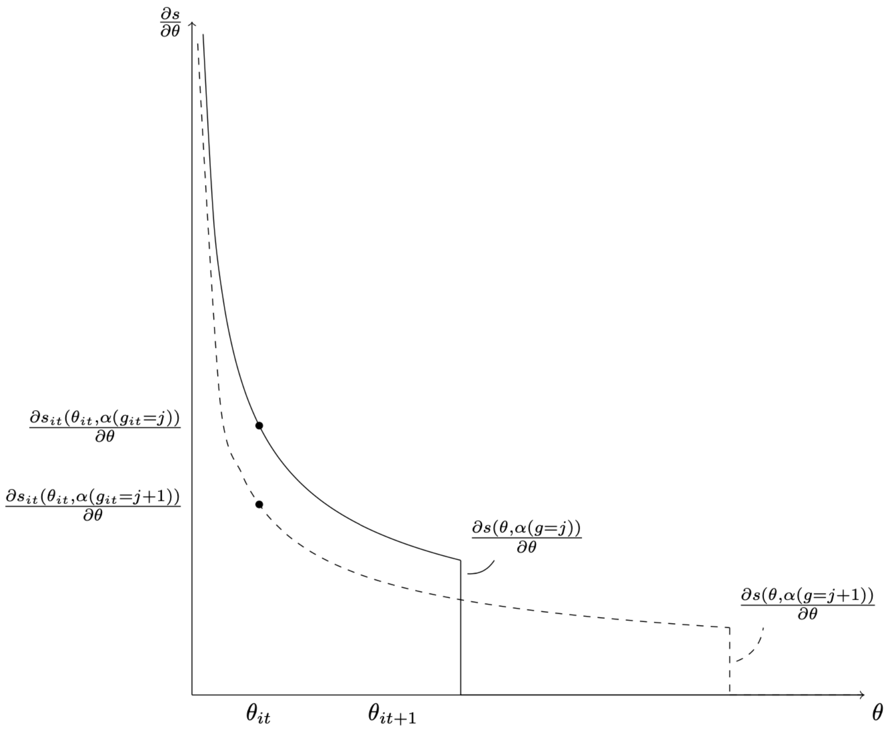


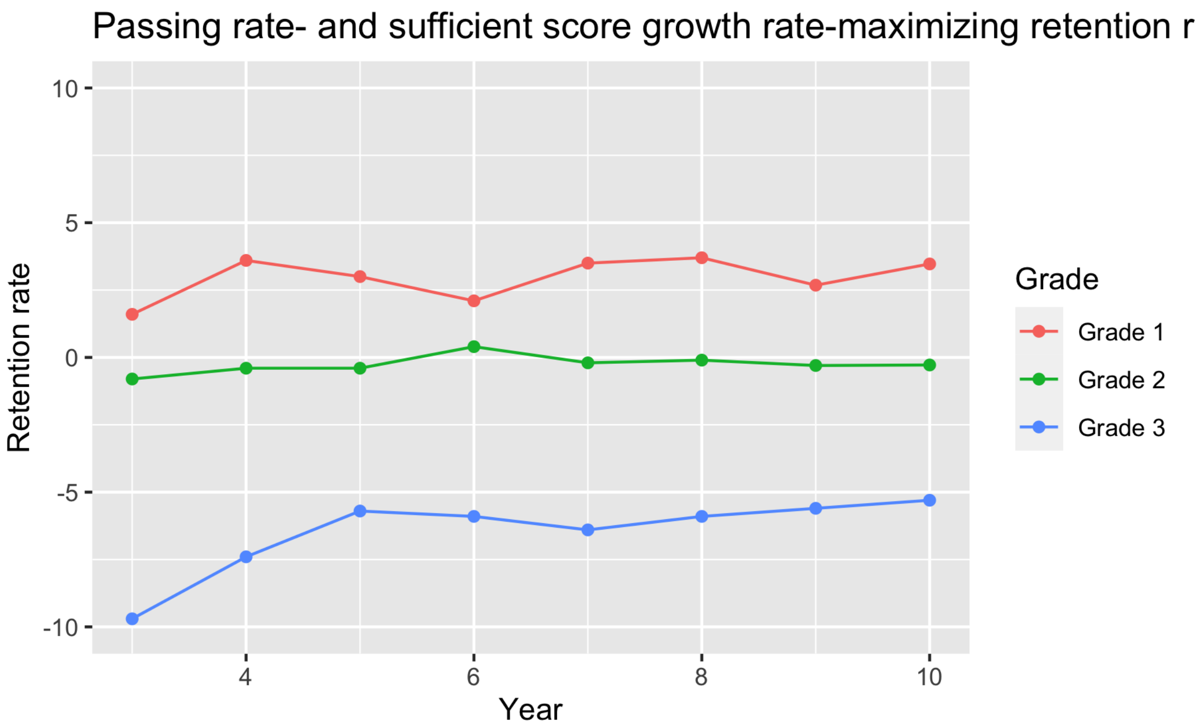
Table 1: Parameter values for simulation

	Mean	S.D.
μ	0	0
ϵ_{it}	0	20
γ_0	35	0
γ_1	1	0
θ_{i0}	50	25[1]
η_{it}	0	10
π	50	0
λ	10	0
Number of students entering grade 1 each year	100	0
Number of years simulated	10	0
Number of simulations	10000	0

Notes:
[1]: Initial θ_{it} draw for students entering school.

The simulation shows that an administrator will use different retention strategies to maximize schoolwide passing rate than he will to maximize both passing rate and sufficient score growth rate. Figure 3 summarizes the results of the simulation. To maximize both passing and sufficient score growth rate, an administrator should set a higher δ_1 on average than he would to maximize passing rate only, and a much lower δ_3. δ_2 should be set similarly as under a passing rate-based system. The optimal δ_1, δ_2, and δ_3 values result in 2.925 percentage-point higher retention rates in the first grade on average (a 37.5% increase), 0.201 percentage-point lower retention rates in grade 2 (a 3% decrease), and 6.485 percentage-point lower retention rates in the third grade on average (a 77% decrease). These results are consistent with the mechanisms shown in Figure 2. Students in grade 1 are expected to satisfy the sufficient growth standard whether promoted or retained.

Figure 2: Passing and growth rate-maximizing retention rates



The simulation further shows that the scores of retained students change in the presence of growth-based accountability. The results show that an administrator retains students with around 5-point higher scores in grade 1, on average, and students with 0.2-point lower scores in grade 2 on average. In the terminal grade, retained students score around 12 points lower on average in the presence of growth-based accountability. In the terminal grade, low-achieving students are relatively likely to pass the growth standard λ, and may achieve the passing threshold π, while high-achieving students are unlikely to pass the growth standard, but more likely to achieve the passing threshold. Growth-based accountability reduces the value of a student achieving the passing threshold only, and as a result, retention rates in the final grade fall.

[input figure summarizing differences in retained students' scores here]

4 Data

I will use restricted student-level data, provided by the TEA and available through the University of Houston's Education Research Center, for this project. Observing student-level test scores and attributes will allow me to perform a student-level analysis of the retention decision, and in particular, an analysis of the way that schools value student test scores in that decision. Since myriad individual attributes factor into the administratorâĂŹs retention decision, an individual-level analysis is critical for understanding how the decision is made. Labadie (2021) analyzes the effects of growth-based accountability on the retention decision, finding that a switch to growth-based accountability reduces retention rates by 18% overall, and that the effect is especially pronounced in the final grade offered by a school. However, the study is unable to analyze which students are less likely to be retained. With individual-level data, I will be able to identify the extent to which administrators retain strategically based on student test scores.

The data I intend to use spans the 1994-95 school year to the 2017-18 school year, and multiple accountability systems. By using data covering such a large period of time, I will be able to analyze the effect of different accountability system components on retention strategy. The data includes observations of the universe of Texas public school students over these 23 years. Though the exact variables that will be available to me are yet to be agreed upon, I intend to use student-level attendance, test score, demographic, special program enrollment status, and behavioral data. I will use these data in conjunction with teacher identifying data, which can be linked to the student data, to account for teacher-specific idiosyncracies in retention. I will also use school ratings data and identifying data for school administrators.

5 Preliminary results and descriptive statistics

Still to come!

6 Empirical strategy

There are a few empirical strategies that might be appropriate for this study, but all require that the data satisfy some identifying assumptions for valid analysis. Without the data to analyze, it is difficult to say which strategies I will employ. One strategy I will attempt is a difference-in-differences approach. I would estimate the following:

$$
\begin{aligned}
y_{sicdt} =& \alpha + \beta_1 \mathbb{1}(t \geq T)_t + \beta_2 \mathbb{1}(g_{sicdt} = g_c^T) && (11) \\
& + \beta_3 \mathbb{1}(t \geq T)_t \times \mathbb{1}(g_{sicdt} = g_c^T) + \gamma_i + \gamma_c + \epsilon_{sicdt} && (12)
\end{aligned}
$$

where y_{iscdt} is equal to one if student s with instructor i in campus c and district d is retained in year t, T represents the year after a change in Texas' accountability system, g_{sicdt} represents the grade of student s, g_c^T represents campus c's final grade, and γ_i and γ_c represent instructor, and campus fixed effects respectively. This strategy is based on the finding of Labadie (2021) that a change to growth-based accountability reduces retention differentially in the final grade offered by schools, as well as the predictions of my conceptual framework. This strategy estimates the difference in the change in within-student probability of retention in terminal grades relative to in non-terminal grades caused by the change in accountability policy. I would use this strategy as a baseline, to demonstrate that the effect documented in Labadie (2021) holds in the Texas context, as well as to find the overall effect of the first accountability policy change, from TAAS to TAKS. To find the role that student scores play in the retention decision, I include score as a third difference:

$$y_{sicdt} = \alpha + \beta_1 \mathbb{1}(t \geq T)_t + \beta_2 \mathbb{1}(g_{sicdt} = g_c^T) + \beta_3 \mathbb{1}(t \geq T)_t \times \mathbb{1}(g_{sicdt} = g_c^T) \quad (13)$$

$$+ \beta_4 S_{sicdt} + \beta_5 \mathbb{1}(t \geq T)_t \times S_{sicdt} + \beta_6 \mathbb{1}(g_{sicdt} = g_c^T) \times S_{sicdt} \quad (14)$$

$$+ \beta_7 \mathbb{1}(t \geq T)_t \times \mathbb{1}(g_{sicdt} = g_c^T) \times S_{sicdt} + \gamma_i + \gamma_c + \epsilon_{sicdt} \quad (15)$$

where S_{sicdt} represents a measure of student performance on the relevant standardized exam. β_7 is the coefficient of interest, representing the marginal effect of an increase in score on retention probability for a student in a terminal grade relative to in a non-terminal one, pre- and post-policy change. I intend to use a measure of distance from the relevant passing standard, or a standardized score, rather than the raw score, since the scores on the tests administered under TAAS, TAKS, and STAAR are not on the same scale. Since score is unlikely to have a linear effect on retention probability, I am likely to use a form of quantile regression. The validity of these methods of identification rely on the retention rates of students in terminal grades and non-terminal grades following parallel trends in the years prior to the change in accountability system.

7 Conclusion

This study proposes to analyze the effects of the adoption of growth-based accountability in Texas on administrator retention strategy. Such a study would be the first to investigate the distributional effects of growth-based accountability. I develop a conceptual framework which proposes a theoretical explanation for why, after the adoption of growth-based accountability, retention rates fall in the terminal grade offered by a school and rise in earlier grades. Administrator incentive to retain students that nearly pass the standardized exam is reduced by the addition of growth-based accountability criteria, and their incentive to retain students at the bottom of the score distribution rises. Students near the middle become less valuable to their scores, particularly in the terminal grade. The results of my study would allow for a better understanding of the effectiveness of adding growth criteria to accountability systems to reduce some of the well-documented unintended consequences of accountability systems, particularly the incentive to focus on high-leverage students.

References

Almond, Douglas, Ajin Lee, and Amy Ellen Schwartz. 2016. "Retention Heterogeneity in New York City Schools."

Brodaty, Thomas O, Robert J Gary-Bobo, and Ana Prieto. 2013. "Does Speed Signal Ability? The Impact of Grade Retention on Wages."

Cockx, Bart, Matteo Picchio, and Stijn Baert. 2017. "Modeling thr Effects of Grade Retention in High School."

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.

Cunha, Flavio, and James J Heckman. 2008. "Formulating , Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *The Journal of Human Resources*, 43(4): 738–782.

Diris, Ron. 2016. "Don't Hold Back? The Effect of Grade Retention on Student Achievement." *Education Finance and Policy*, 11(3): 225–50.

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.

Figlio, David N, and Umut Özek. 2019. "An Extra Year to Learn English? Early Grade Retention and the Human Capital Development of English Learners."

Fruehwirth, Jane Cooley, Salvador Navarro, and Yuya Takahashi. 2016. "How the Timing of Grade Retention Affects Outcomes : Identification and Estimation of Time-Varying Treatment Effects." *Jornal of Labor Economics*, 34(4): 979–1021.

Gary-Bobo, Robert J., Marion Goussé, and Jean-Marc Robin. 2016. "Grade retention and unobserved heterogeneity." *Quantitative Economics*, 7(3): 781–820.

Greene, Jay P., and Marcus A. Winters. 2007. "Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy." *Education Finance and Policy*, 2(4): 319–340.

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.

Labadie, William. 2021. "Growth-based Accountability Rules and Grade Retention Practices."

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

Radcli, Jennifer, and Keri Blakinger. 2017. "Held Back : Some Houston-area schools have high retention rates for elementary students."

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

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.