Research brief: How did STAAR change Texas retention practices?

(Latest version can be found here)

William Labadie

October 26, 2020

Abstract

The switch from TAKS to STAAR changed the way in which Texas schools are evaluated by the Texas Education Agency (TEA), and in so doing, changed the objectives of school administrators in the state. This study explores the way that the change to administrator objectives affected retention practices in Texas public schools, and whether there were any long-term effects to students' future exam scores.

1 Background

Public schools in Texas are evaluated by the Texas Education Agency (TEA) based on student performance on standardized statewide tests, and have been since the 1993-94 school year. A large body of research has demonstrated that administrators are sensitive to the ratings assigned to their schools, and that they use numerous tools to achieve as high a rating as possible based on the relevant school rating criteria. [1] If an administrator considers the effect on his school's rating when choosing whether or not to retain a student in-grade, his decision will depend on the ratings criteria his school faces. Through retention, an administrator has a degree of control over the difficulty of test that a given student faces, and is able to affect the composition of the pool of test-takers in his school.[2] In the final grade offered by the school, the administrator is able to remove test-takers that may harm his school's rating by promoting them, or include test-takers that may improve the rating by retaining them.

Though retention is not a focus of the STAAR accountability system, it is a worthwhile intervention to study in this context. While the literature studying the effects of retention on retained students has found generally mixed results, researchers have found a range of long-term effects, including increased risk of future dropout (Manacorda, 2012) and non-completion of high school (Jacob and Lefgren, 2009), increased likelihood to be convicted of a crime by 25 (Eren, Lovenheim and Mocan, 2018), and decreased future wages (Brodaty, Gary-Bobo and Prieto, 2013). At the same time, some researchers have found that retention improves in student achievement on standardized tests, particularly after retention in elementary school (Jacob and Lefgren, 2004; Schwerdt, West and Winters, 2017; Nunes, Reis and Seabra, 2018). Retention clearly affects both a student's short-term achievement and longer-term development. These two forces may be salient to Texas administrators given the change in ratings criteria brought about by the adoption of the STAAR system.

The STAAR test was first administrated in 2012, and scores were first used to evaluate schools in 2013. The STAAR-based accountability system replaced the existing accountability system, which was based on the Texas Assessment of Knowledge and Skills (TAKS). Under the TAKS-based system, schools were evaluated based on student achievement on TAKS, attendance rates, dropout rates for grades 7-12, and college readiness. Under the STAAR-based system, schools are evaluated based on student achievement on STAAR, student progress on STAAR, achievement on STAAR by the lowest-performing two racial/ethnic student groups relative to the achievement of the rest of the student body, and postsecondary readiness (Texas Education Agency, 2017).[3]Under both systems, students who failed both the math and reading test in 5th and 8th grade were not allowed to move on to the next grade. With the inclusion of student progress (as measured by a formula based on score improvement on the STAAR from one year to the next), an administrator considering retention as a tool

[1]Figlio and Winicki (2005) and Reback (2008) find targeted efforts to improve test scores when schools are rated based on student body passing rate.

[2]Cullen and Reback (2006) find that administrators in Texas strategically exempt students from the exam to improve school overall passing rate.

[3]Student progress is measured within racial/ethnic and academic groups, such as special education and English language proficiency level.

for improving her school's rating must consider both whether retention will make a student more likely to pass the exam (boosting the student achievement score of the school) and whether it will make the student more likely to demonstrate sufficient score growth (boosting the student progress score of the school).

2 Research Questions

In this study, I exploit the unique nature of the final grade offered by a school, which allows an administrator to remove a student from the test-taking pool via promotion, to answer:

- How did the switch to STAAR affect retention practices in Texas public schools?

I then focus on differences in exposure to the changes caused by STAAR across cohorts to see:

- How do more-affected cohorts perform over their school careers relative to less-affected cohorts?

3 Empirical Analyses

Question 1: How did the switch to STAAR affect retention practices in Texas public schools?

Given the effects retention can have on students, understanding the relationship between administrators' incentives and their retention practices could be important in implementing new accountability standards. This information may be useful to legislators and regulators as retention is not an outcome that is explicitly meant to be affected by the STAAR accountability system.

The primary data source used to answer this question is a school-grade level panel provided by TEA. These data include retention rates and enrollment totals for various demographic groups, and test scores and passing rates for reading and math exams, going back to the 2003-04 school year (the first year of TAKS). The sample is limited to 3rd to 9th grade; since 5th and 8th grade are both common terminal grades offered by schools and have their own retention-related restriction, much of the attention of this study will be paid to those grades.

Table 1 presents retention rates before and after the adoption of STAAR for all, male, female, white, Black, and Hispanic students. These are the largest demographics on average, and are the most likely to be reported by schools.[4] Table 1 shows that retention rates fall for every group presented under STAAR, both in terminal and non-terminal grades. Retention rates in terminal grades are lower on average than those in non-terminal grades. There are no clear differences across groups, so I focus on all students for the remainder of the study. Since a high proportion of schools end in 5th and 8th grade, I present retention rates by grade for terminal and non-terminal grades in Table 2. For each grade, retention rates are lower when the grade is terminal. The reduction in retention rates associated with the switch to STAAR seem steepest in 5th-8th grade, regardless of terminality; terminal grades seem to face similar drops as non-terminal grades.

[4]While other demographic groups matter to schools' ratings, there is a relatively high degree of missingness in the retention data for many of these groups, and so I omit these groups from this study.

Table 1: Retention rate, by STAAR exposure and terminality

	Non-terminal grades		Terminal grades	
	Pre-STAAR	Post-STAAR	Pre-STAAR	Post-STAAR
All students	4.188	3.133	2.123	1.160
	(9.767)	(8.123)	(8.583)	(4.837)
Male students	4.612	3.486	2.067	1.241
	(10.30)	(8.686)	(7.983)	(4.870)
Female students	3.206	2.477	1.590	0.923
	(8.783)	(7.347)	(6.138)	(3.672)
White students	3.035	2.634	1.334	0.981
	(10.49)	(9.643)	(7.342)	(5.334)
Black students	3.387	3.052	1.527	1.112
	(11.12)	(10.74)	(7.264)	(5.768)
Hispanic students	4.291	3.259	2.013	1.174
	(10.59)	(8.900)	(7.571)	(4.895)
N	212929	160739	25816	18216

mean coefficients; sd in parentheses

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

Table 2: Retention rates, by STAAR exposure and terminality

	Non-terminal grades		Terminal grades	
	Pre-STAAR	Post-STAAR	Pre-STAAR	Post-STAAR
Grade 3	2.491	1.926	1.659	1.378
	(4.364)	(3.567)	(2.167)	(1.749)
N	28732	21846	658	429
Grade 4	1.545	1.152	0.988	0.863
	(4.356)	(3.234)	(2.506)	(1.493)
N	26931	20425	2380	1763
Grade 5	2.336	1.533	1.483	0.866
	(6.493)	(4.583)	(2.961)	(1.866)
N	12482	9286	15171	11667
Grade 6	2.379	1.389	0.867	0.632
	(9.479)	(6.140)	(5.735)	(4.483)
N	13052	10116	4791	2672
Grade 7	3.048	1.901	17.384	2.548
	(9.912)	(7.662)	(36.692)	(13.637)
N	12636	11512	85	54
Grade 8	3.023	1.810	3.146	1.799
	(9.584)	(7.083)	(10.909)	(6.740)
N	14793	11025	1199	666

mean coefficients; sd in parentheses.

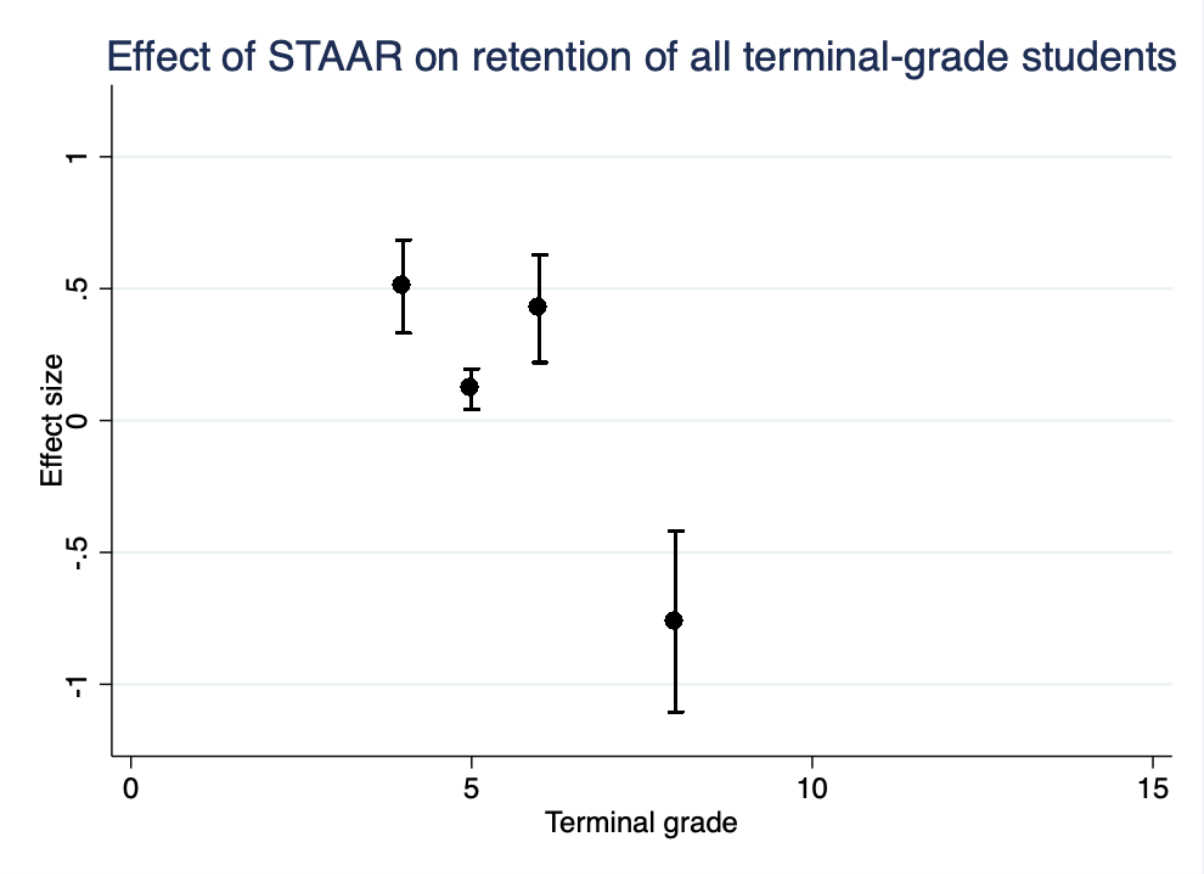
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Using a fixed effects difference-in-differences regression strategy, I compare the within-school retention rate in the terminal grade relative to all other grades post-STAAR relative to the same difference pre-STAAR.[5] The results of this regression for all students can be found in Appendix A, Table A.1 (the same regression, run separately for each demographic group, can be found in Appendix A as well). For this regression to produce a causal policy effect estimate, there cannot be differential trends in retention rate between terminal and non-terminal grades prior to the adoption of STAAR in 2013. Unfortunately, this requirement only reliably holds in 5th grade - as shown in Table 2, this is by far the most frequent terminal grade in this sample - and in 4th grade. The estimated effects for all other grades should not be interpreted as causal.

Figure 1 presents the estimated effect of STAAR on retention practices in Texas. I find that STAAR increases the average retention rate of terminal-grade 5th graders by 0.5 percentage points, or 40 percent. This result holds across demographic groups (Check Appendix A.3 for similar graphs for each demographic group). This result suggests that administrators believe retaining 5th graders, rather than promoting them out of school, will improve their ratings after the criteria change.

[5]Details on this empirical strategy can be found in Appendix B.

Figure 1: Regression results, all students



Notes: Based on the regression results reported in Table A.1. 95% confidence intervals are plotted.

3.1 Question 2: How do more-affected cohorts perform over their school careers relative to less-affected cohorts?

It's important to understand the long-term effects of the change in retention practices caused by STAAR. Though a change to retention is an unintended consequence of the implementation of STAAR, it's worthwhile to examine the long-term effects of the increased retention on student outcomes to attempt to understand whether this harms or benefits students or schools.

Table 3 presents the retention rates by grade for cohorts more likely to be affected by the change in retention practice - those entering 5th grade in 2013 or later - and cohorts less likely to be affected by the change - those who entered 5th grade by 2012. The table shows the more-affected cohort ("Young" in the table) retained at a higher rate in 5th grade than the less-affected cohort ("Old"), and at a lower rate in 6th, 7th, and 8th grade. These trends hold for all demographic groups. Table 4 presents the average math STAAR scores by grade for cohorts more likely to be affected by the change in retention practices and cohorts less likely to be affected by the change in retention practices. Unfortunately, since the less-affected cohort was in 5th grade or higher in 2012, I cannot compare the 3rd and 4th grade scores of the groups. This is unfortunate, since 5th grade is the most clearly affected grade based on the results reported in Figure 1, and comparing the two cohorts' scores before and after passing through 5th grade would have been informative. The table shows that the younger cohort scores better than the older cohort across all demographic groups.

Table 3: Retention rate by exposure

	Grade 5		Grade 6		Grade 7		Grade 8	
	Old	Young	Old	Young	Old	Young	Old	Young
All students	0.742	1.162	1.640	1.148	2.334	1.620	2.049	1.464
	(3.910)	(3.370)	(7.684)	(5.563)	(8.499)	(6.972)	(7.611)	(6.410)
Male students	0.820	1.267	1.789	1.309	2.695	1.738	2.065	1.423
	(3.820)	(3.999)	(7.724)	(5.733)	(9.249)	(6.829)	(7.764)	(6.296)
Female students	0.532	1.009	1.094	0.767	1.459	1.153	1.679	1.252
	(2.926)	(3.332)	(6.592)	(4.931)	(7.370)	(6.613)	(7.522)	(6.991)
White students	0.771	0.984	1.123	0.836	1.888	1.214	1.562	1.146
	(5.552)	(5.549)	(6.436)	(5.094)	(9.612)	(6.939)	(7.931)	(7.428)
Black students	0.628	1.195	1.295	0.926	1.764	1.311	1.575	1.157
	(4.661)	(6.160)	(7.145)	(5.861)	(8.446)	(7.642)	(8.315)	(7.306)
Hispanic students	0.744	1.258	1.589	1.149	2.246	1.584	2.040	1.355
	(3.788)	(4.242)	(8.039)	(6.323)	(8.953)	(7.459)	(8.152)	(6.728)
N	4086	20953	5131	10241	6912	6944	9343	4685

mean coefficients; sd in parentheses

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

Table 4: STAAR achievement by exposure

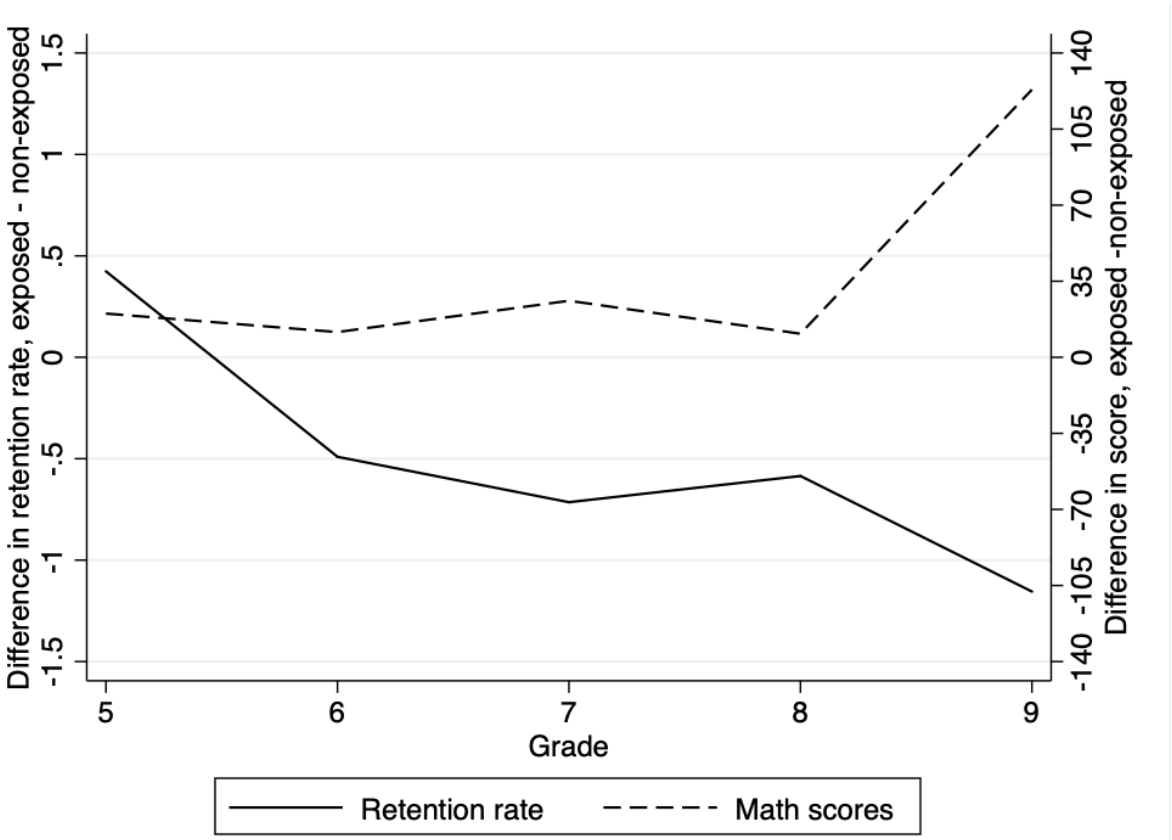
	Grade 5		Grade 6		Grade 7		Grade 8	
	Old	Young	Old	Young	Old	Young	Old	Young
All students	1578.6	1598.9	1611.3	1623.0	1618.4	1644.5	1654.3	1665.2
	(60.26)	(65.73)	(68.76)	(68.65)	(60.04)	(68.26)	(60.20)	(68.42)
Male students	1578.8	1599.6	1612.9	1625.3	1619.2	1644.9	1655.5	1660.5
	(62.18)	(68.72)	(70.79)	(70.13)	(61.28)	(68.72)	(60.40)	(68.38)
Female students	1580.5	1599.9	1612.2	1621.6	1620.5	1647.3	1657.7	1675.4
	(61.28)	(65.85)	(68.40)	(66.94)	(59.36)	(66.85)	(59.46)	(66.24)
White students	1611.3	1630.8	1646.2	1653.2	1647.0	1671.8	1679.7	1690.1
	(63.71)	(68.17)	(69.91)	(69.68)	(57.81)	(66.68)	(58.68)	(65.83)
Black students	1539.5	1554.5	1567.0	1584.8	1584.1	1605.8	1629.1	1639.0
	(59.75)	(65.39)	(61.67)	(60.12)	(52.26)	(59.54)	(51.32)	(59.38)
Hispanic students	1569.2	1590.0	1597.8	1609.4	1607.0	1632.9	1647.1	1659.1
	(53.10)	(59.18)	(61.11)	(59.12)	(52.52)	(58.38)	(52.88)	(61.81)
N	3946	20347	4701	9469	6087	6251	8052	4080

mean coefficients; sd in parentheses

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

Figure 2 shows the difference between the math achievement of the exposed cohort and the less-exposed cohort from 5th to 9th grade.[6] The figure shows that the more-exposed cohort is retained at a higher rate than the less-exposed cohort in 5th grade, and at a lower rate in 6th-8th grade. At the same time, the more-exposed cohort score consistently higher on the math STAAR than the less-exposed cohort. It's possible that the policy change causes administrators to retain students earlier in their careers, and in so doing actually improving their achievement.

Figure 2: More retention for the exposed - and higher scores



Notes: Data is from Texas, spanning the 2011-12 to 2016-17 school years.

[6]High school STAAR requirements are course-based rather than level-based; the score used here is that of Algebra I.

4 Discussion and Recommendations

In this study, I present some of the effects of Texas' adoption of the STAAR test. I explore retention with a focus on the terminal grade of a school, where an administrator's retention decision may have the most salience with respect to the school's rating. Given that retention is not an area addressed by STAAR, it is important to understand how the retention decision is affected by the policy change.

The analyses show that administrators do change their retention practices in response to the policy change. Specifically, I find that the policy change causes a 40% increase in retention in 5th grade, when 5th grade is the final grade offered by a school. The presence of a differential effect in the terminal grade offered by schools suggests that administrators in Texas do attempt to use retention to improve their school ratings, and that they adjust their strategy based on the criteria by which their schools are rated. If TEA does not wish to affect retention practices, the current formulation of the student progress measure used to evaluate schools may be worth re-evaluating.

The results of this study also show that cohorts entering 5th grade after the adoption of STAAR have higher math achievement scores from 5th-9th grade than cohorts exposed to STAAR in the 6th grade or later. It is possible that the additional criteria schools must satisfy under STAAR lead administrators to target retention more appropriately than under TAKS. Because an administrator has more flexibility in the ratings she must satisfy, a retained student that fails an exam may not harm the school's overall rating if her scores increase by a sufficient amount relative to the previous year's; administrators' incentives may lead them to retain students they believe would be best served by retention under STAAR, who they would not retain under TAKS. Scores are an important out come for schools and administrators, but less so for students. Further investigation of the developmental implications of the extra retention found by this study should inform any analysis of this unintended change in administrator behavior.

References

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

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System."

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.

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." *Journal of Public Economics*, 92: 1394–1415.

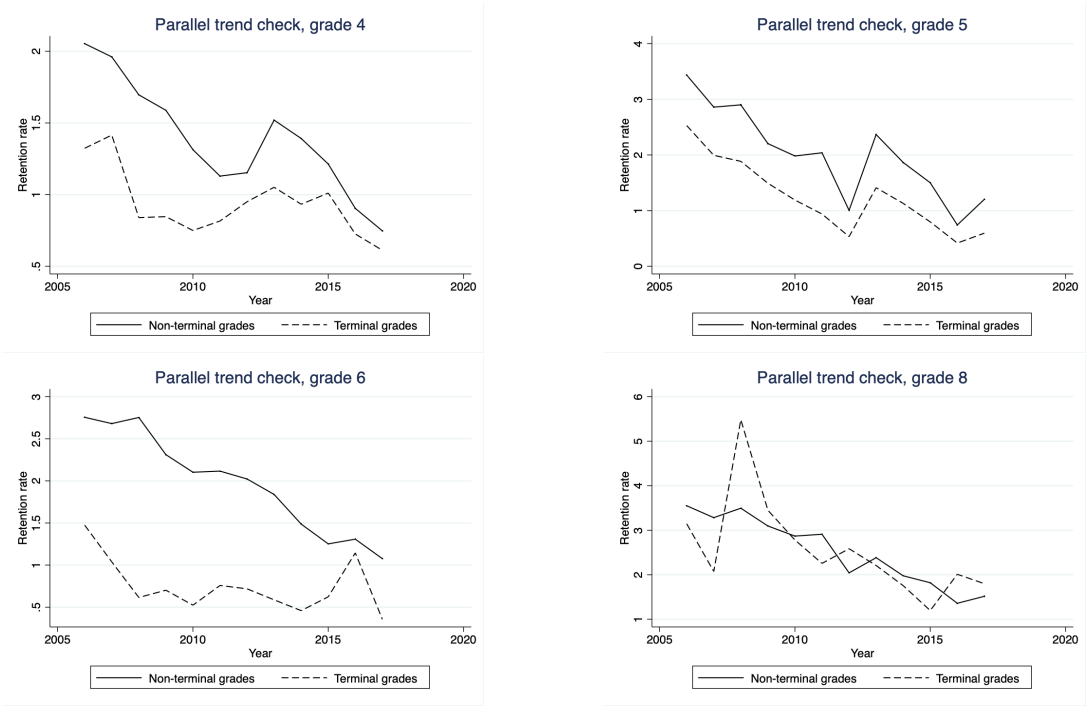
Schwerdt, Guido, Martin R. West, and Marcus A. Winters. 2017. "The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida." *Journal of Public Economics*, 152: 154–169.

Texas Education Agency. 2017. "2017 Accountability Manual."

A Additional tables & figures

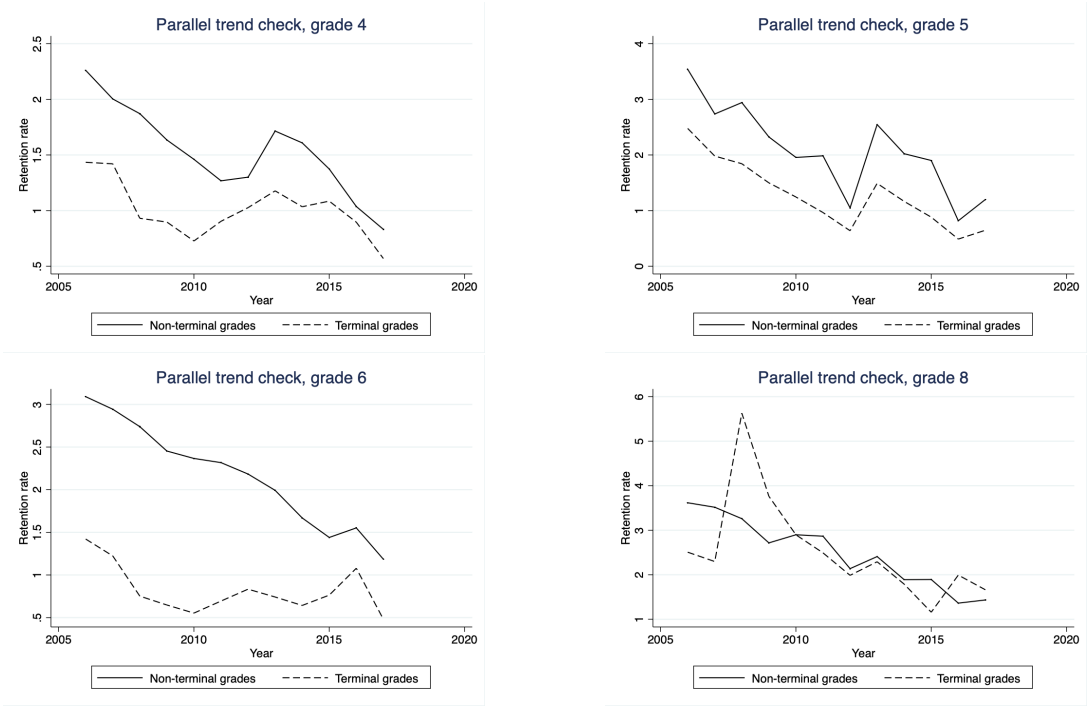
A.1 Parallel trend checks

Figure A.1: Parallel trend checks, all students



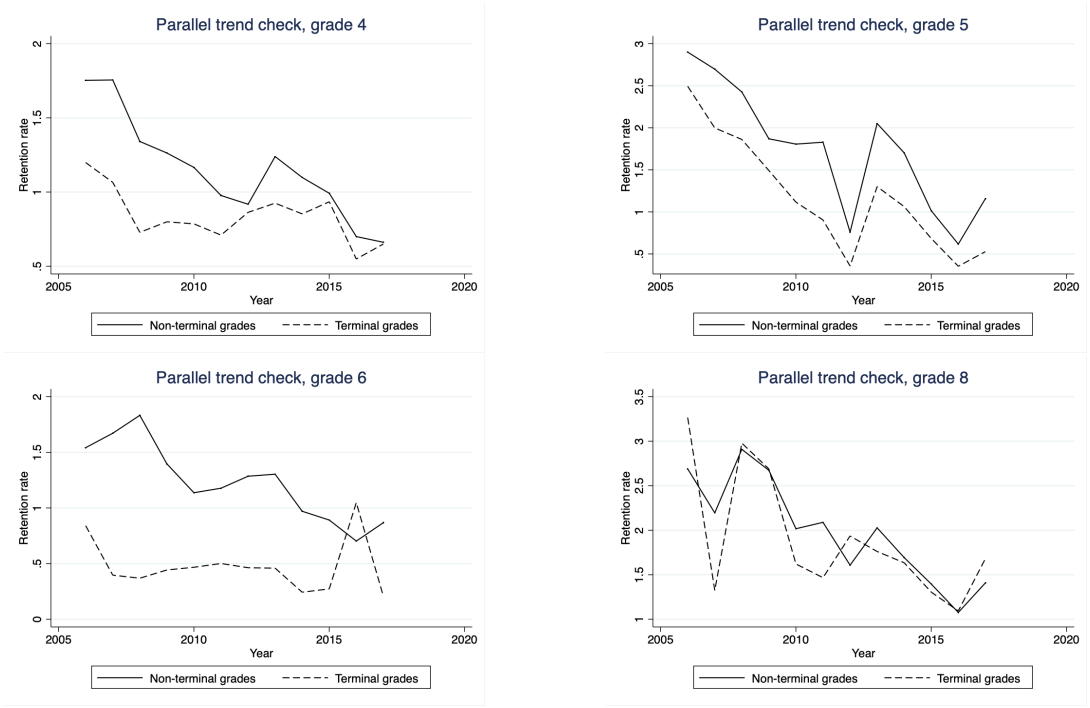
Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

Figure A.2: Parallel trend checks, male students



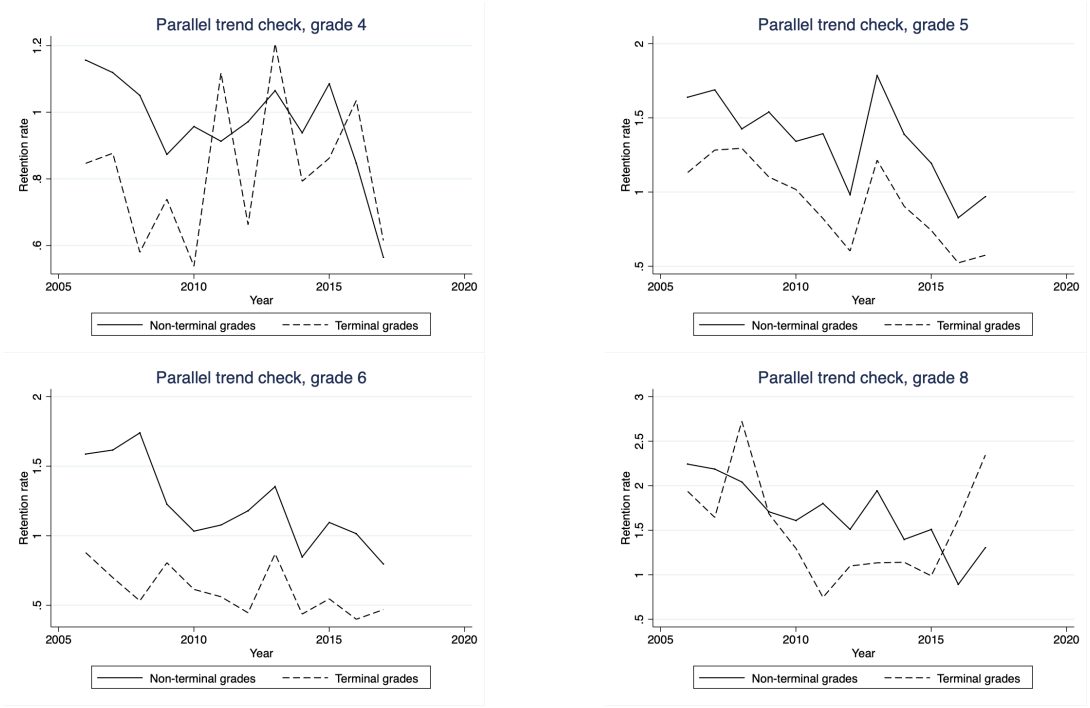
Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

Figure A.3: Parallel trend checks, female students



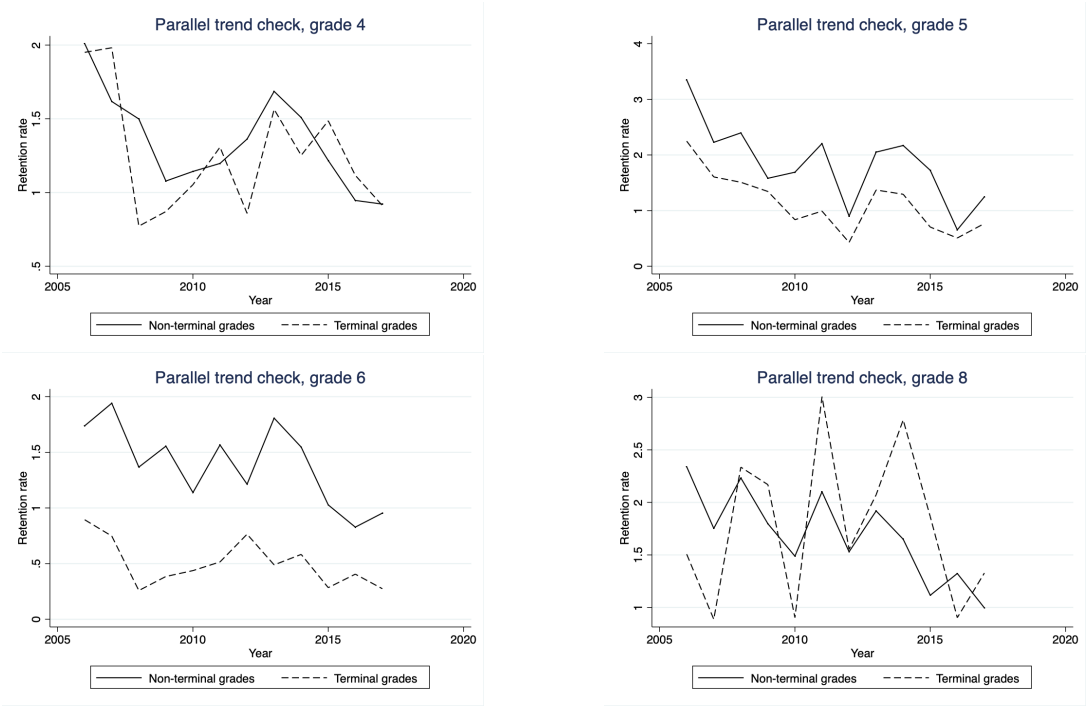
Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

Figure A.4: Parallel trend checks, white students



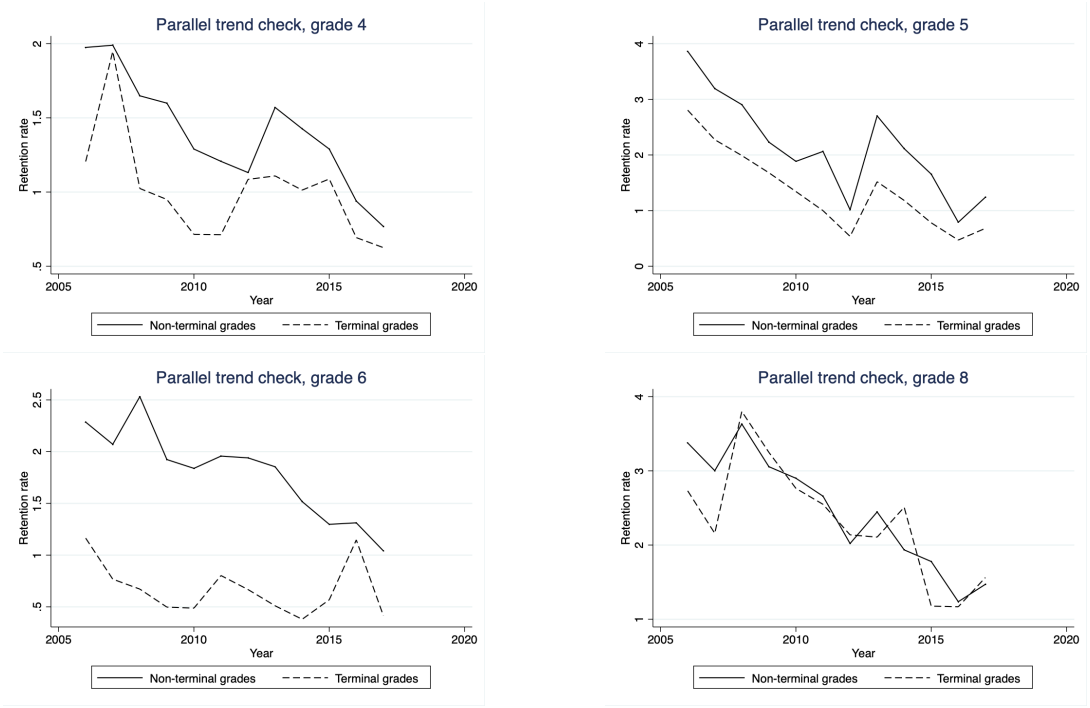
Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

Figure A.5: Parallel trend checks, Black students



Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

Figure A.6: Parallel trend checks, Hispanic students



Notes: Based on the regression results reported in Table tk. 95% confidence intervals are plotted.

A.2 DID regression results

Table A.1: Effect of STAAR on retention of all students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.508**	0.119	0.424*	-0.761*
	(0.177)	(0.077)	(0.204)	(0.343)
$\mathbb{1}$(Terminal grade)	-2.844***	-1.685***	-1.858***	-0.138
	(0.124)	(0.057)	(0.118)	(0.282)
$\mathbb{1}$(Post-STAAR)	-0.547***	-0.634***	-0.547***	-0.379**
	(0.096)	(0.054)	(0.064)	(0.121)
Constant	3.814***	3.155***	2.774***	2.926***
	(0.064)	(0.044)	(0.043)	(0.084)
N	21288	164059	74358	15174

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

Table A.2: Effect of STAAR on retention of male students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.671***	0.316***	0.611***	-0.886*
	(0.199)	(0.088)	(0.182)	(0.371)
$\mathbb{1}$(Terminal grade)	-3.355***	-2.205***	-2.260***	-0.500
	(0.143)	(0.065)	(0.112)	(0.304)
$\mathbb{1}$(Post-STAAR)	-0.695***	-0.778***	-0.636***	-0.277
	(0.107)	(0.062)	(0.074)	(0.147)
Constant	4.393***	3.692***	3.225***	3.309***
	(0.072)	(0.050)	(0.049)	(0.096)
N	21283	164010	74288	15108

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

Table A.3: Effect of STAAR on retention of female students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.352*	-0.102	0.449*	-0.506
	(0.160)	(0.070)	(0.186)	(0.308)
$\mathbb{1}$(Terminal grade)	-2.309***	-1.128***	-1.717***	-0.114
	(0.100)	(0.052)	(0.086)	(0.258)
$\mathbb{1}$(Post-STAAR)	-0.369***	-0.470***	-0.444***	-0.121
	(0.093)	(0.047)	(0.058)	(0.122)
Constant	3.157***	2.554***	2.258***	2.173***
	(0.061)	(0.038)	(0.039)	(0.075)
N	21282	163842	74230	14985

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

Table A.4: Effect of STAAR on retention of white students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.478*	-0.057	0.078	-0.101
	(0.206)	(0.095)	(0.129)	(0.412)
$\mathbb{1}$(Terminal grade)	-2.529***	-1.320***	-1.389***	-0.244
	(0.144)	(0.067)	(0.098)	(0.281)
$\mathbb{1}$(Post-STAAR)	-0.297*	-0.115*	-0.123	0.051
	(0.139)	(0.058)	(0.066)	(0.148)
Constant	3.312***	2.371***	2.098***	1.935***
	(0.073)	(0.037)	(0.039)	(0.086)
N	20909	151261	69174	13715

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

Table A.5: Effect of STAAR on retention of Black students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.430	0.104	0.303*	-0.689
	(0.304)	(0.118)	(0.146)	(0.480)
$\mathbb{1}$(Terminal grade)	-2.880***	-1.682***	-2.029***	-0.500
	(0.199)	(0.081)	(0.109)	(0.261)
$\mathbb{1}$(Post-STAAR)	-0.484**	-0.607***	-0.545***	0.512*
	(0.176)	(0.084)	(0.095)	(0.241)
Constant	4.160***	3.070***	2.721***	2.000***
	(0.121)	(0.062)	(0.060)	(0.116)
N	19750	142623	65742	11897

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

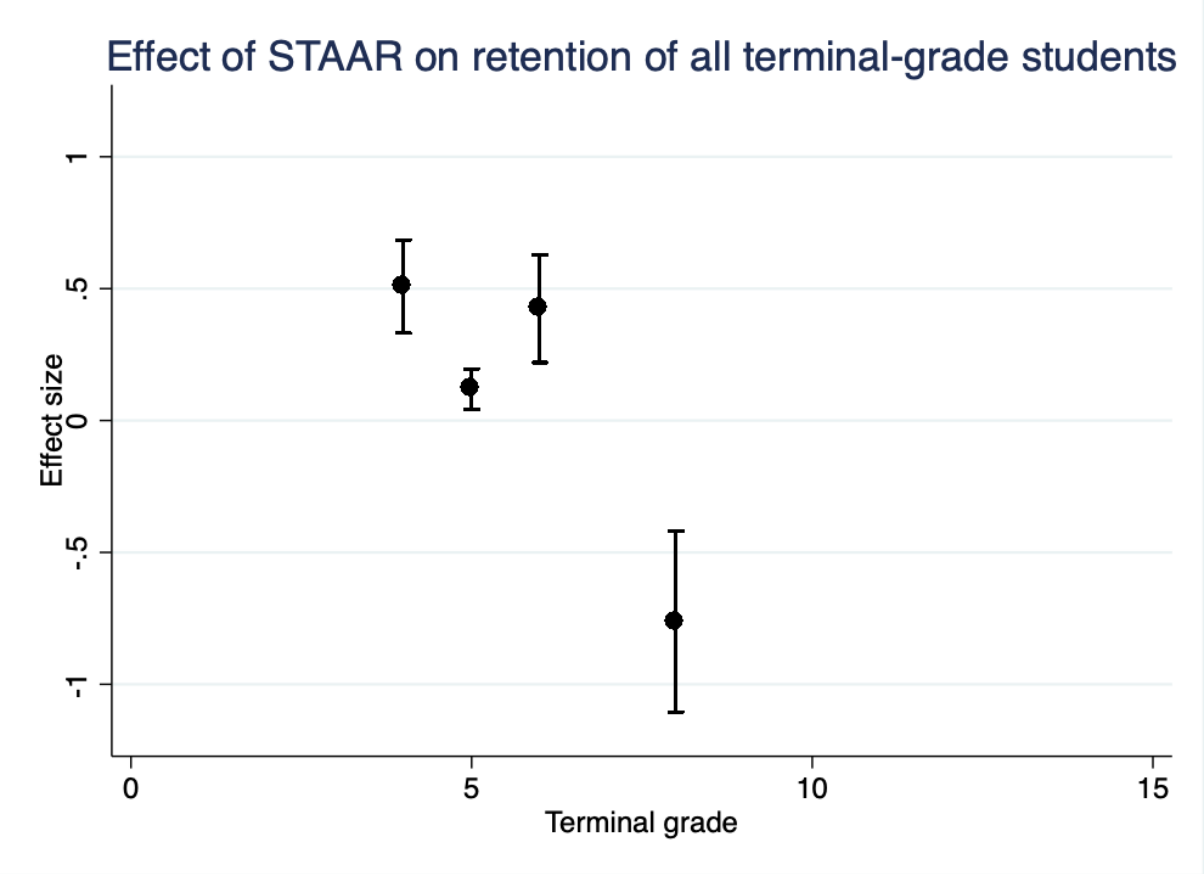
Table A.6: Effect of STAAR on retention of Hispanic students, by grade

	Terminal grade is:			
	Grade 4	Grade 5	Grade 6	Grade 8
$\mathbb{1}$(Terminal grade) $\times$ $\mathbb{1}$(Post-STAAR)	0.588**	0.159	0.610**	-0.926*
	(0.185)	(0.083)	(0.199)	(0.363)
$\mathbb{1}$(Terminal grade)	-3.182***	-1.827***	-2.345***	-0.387
	(0.130)	(0.059)	(0.095)	(0.276)
$\mathbb{1}$(Post-STAAR)	-0.691***	-0.790***	-0.690***	-0.310
	(0.119)	(0.057)	(0.073)	(0.183)
Constant	4.248***	3.461***	3.080***	3.061***
	(0.079)	(0.045)	(0.051)	(0.106)
N	21242	163685	73847	14631

Notes: Standard errors, clustered at the district-year level, are reported in parentheses. *, **, and *** denote statistical significance at the 90, 95, and 99% levels respectively. All regressions include school-level fixed effects.

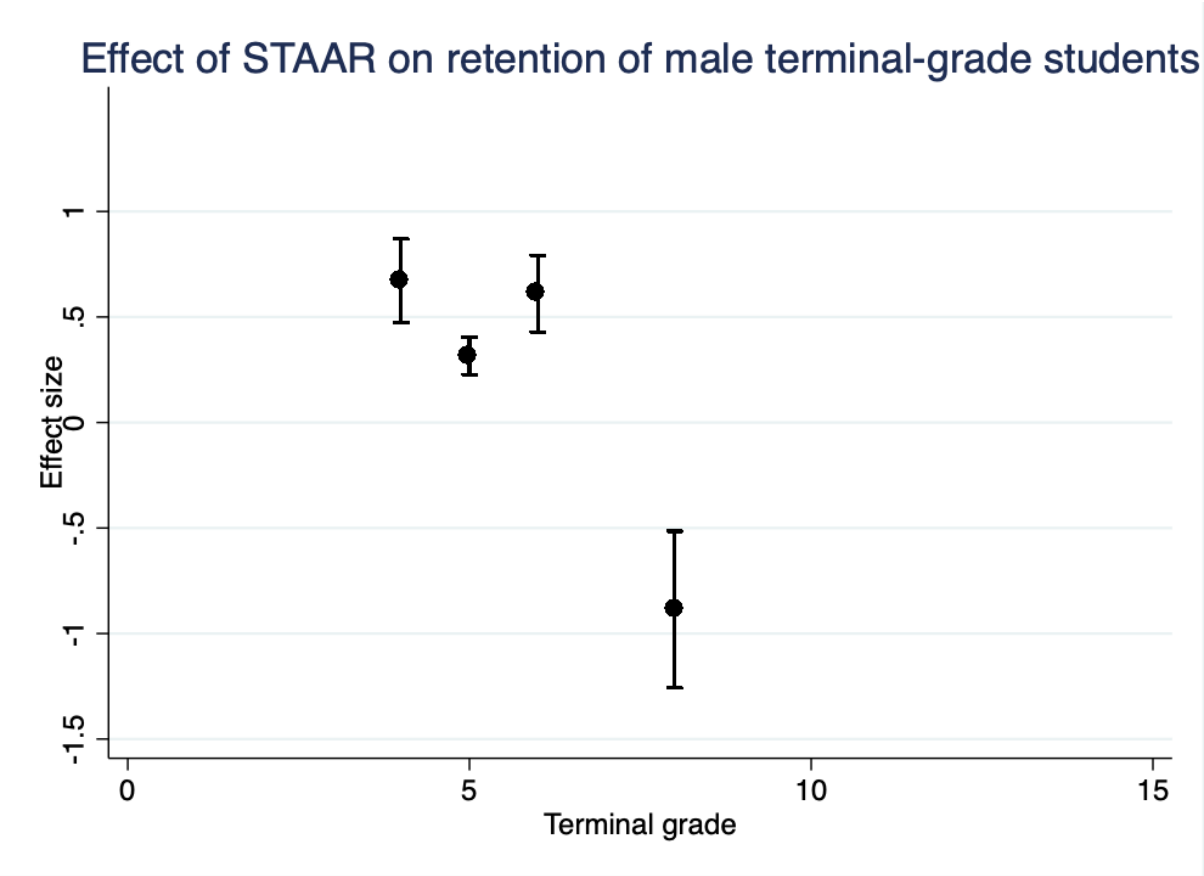
A.3 DID effect plots

Figure A.1: Regression results, all students



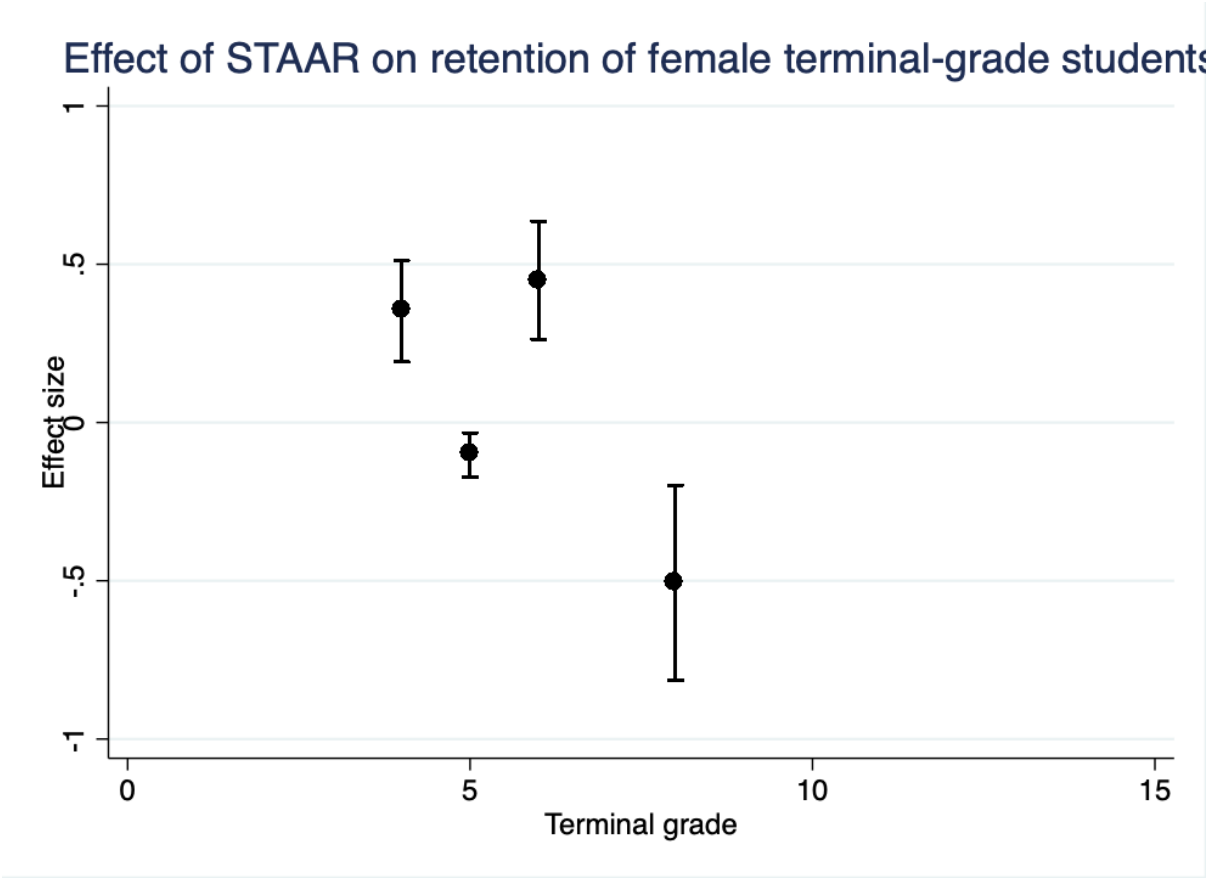
Notes: Based on the regression results reported in Table A.1. 95% confidence intervals are plotted.

Figure A.2: Regression results, male students



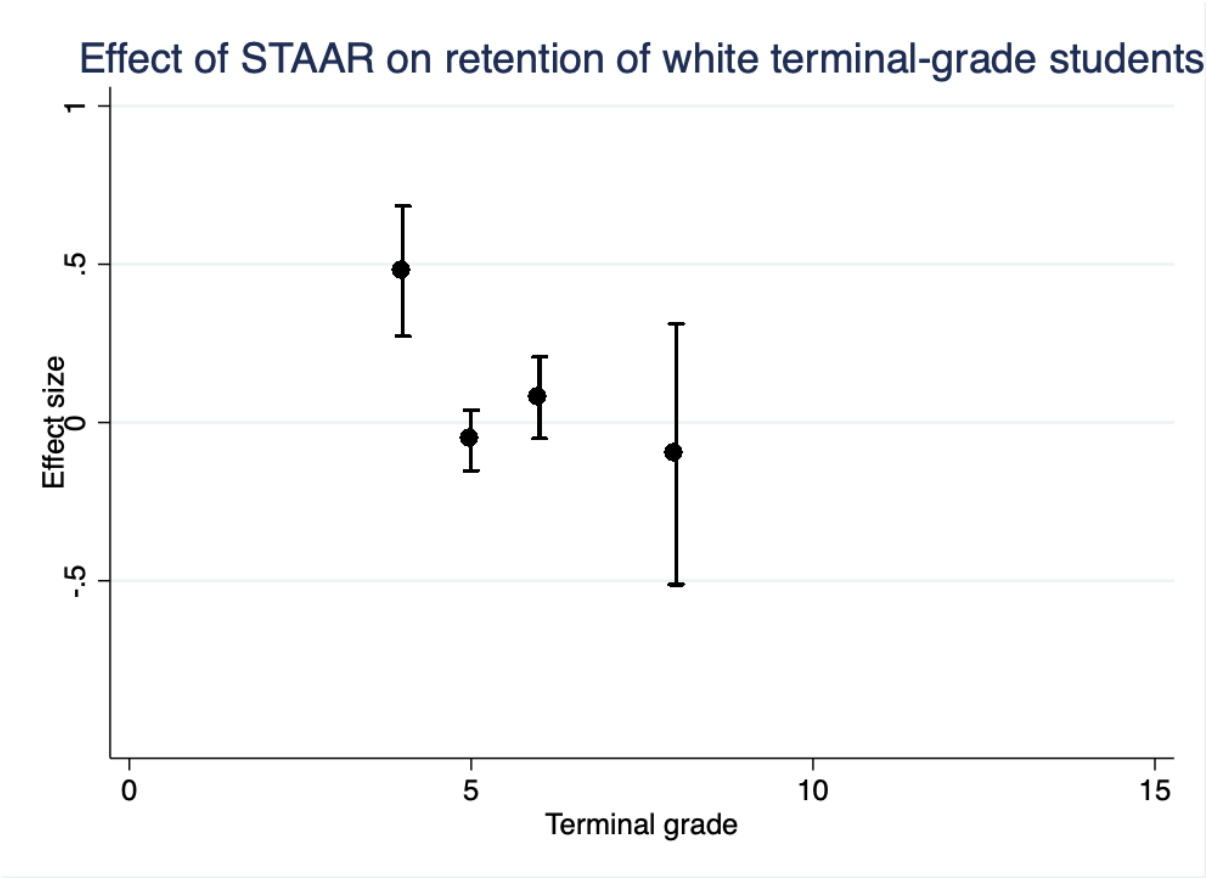
Notes: Based on the regression results reported in Table A.2. 95% confidence intervals are plotted.

Figure A.3: Regression results, female students



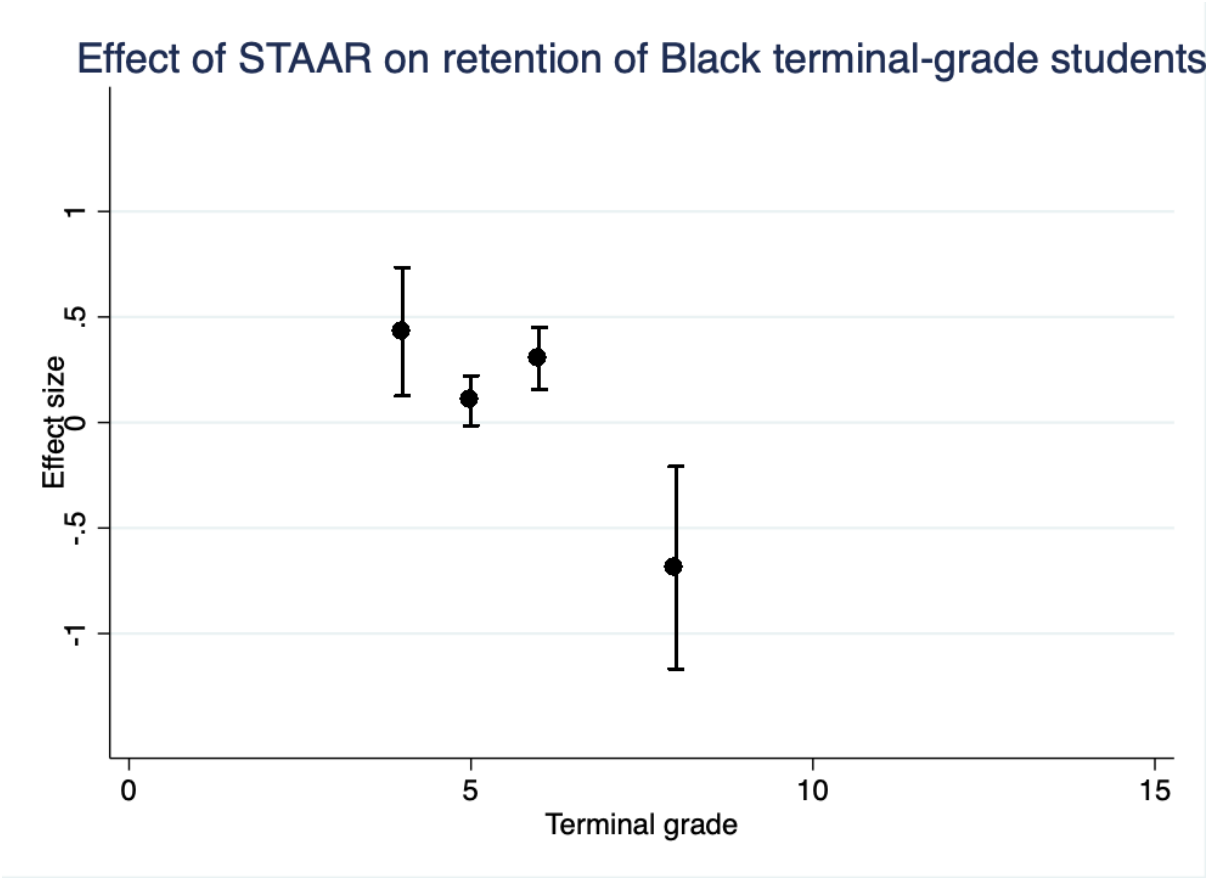
Notes: Based on the regression results reported in Table A.3. 95% confidence intervals are plotted.

Figure A.4: Regression results, white students



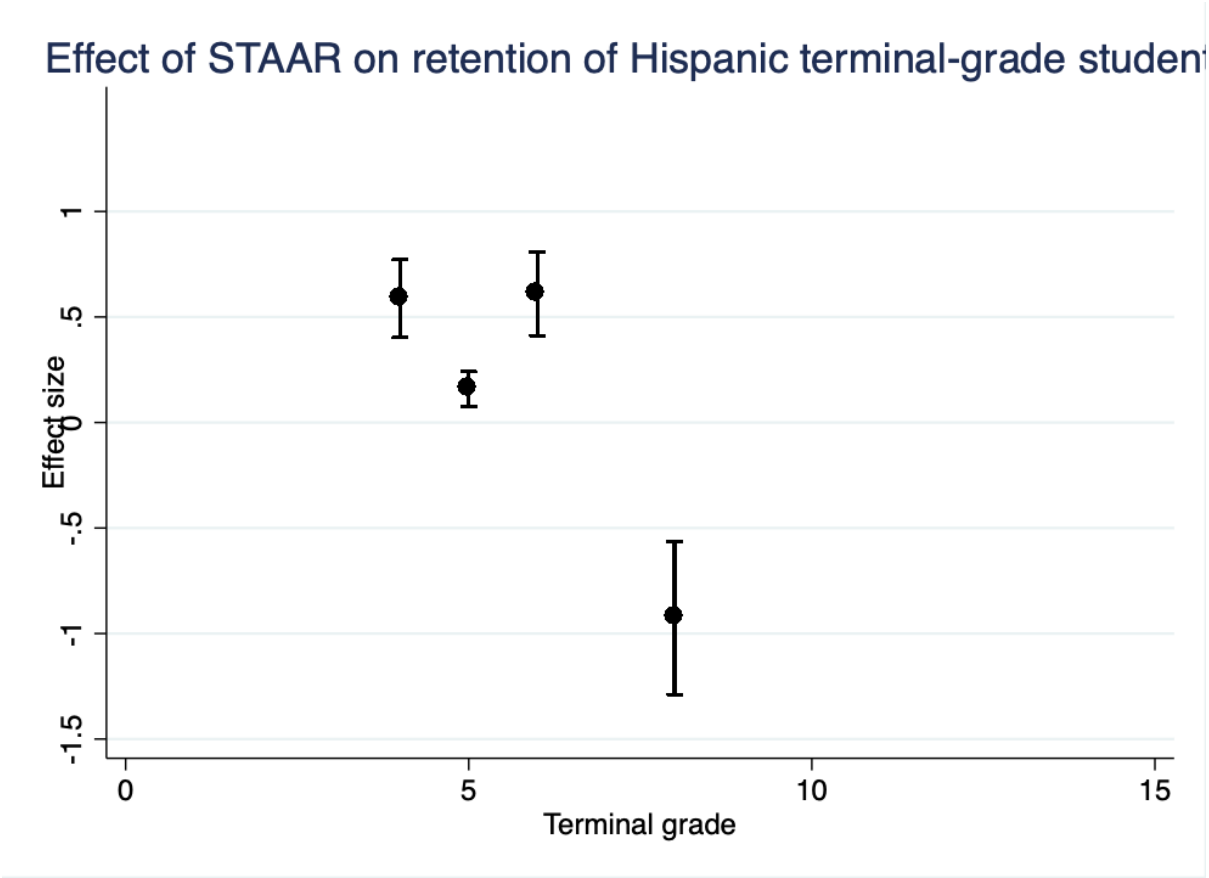
Notes: Based on the regression results reported in Table A.4. 95% confidence intervals are plotted.

Figure A.5: Regression results, Black students



Notes: Based on the regression results reported in Table A.5. 95% confidence intervals are plotted.

Figure A.6: Regression results, Hispanic students



Notes: Based on the regression results reported in Table A.6. 95% confidence intervals are plotted.

B Empirical methodology

To test the effect of the adoption of STAAR on retention practices in the last grade offered by a school, I estimate the following equation separately for 4th-8th grade. I estimate a regression of the following form:

$$\begin{aligned} r_{gcdt} =& \alpha + \beta_1 \mathbb{1}(t \geq 2013)_t \times \mathbb{1}(g = G_{ct}^T) + \beta_2 \mathbb{1}(t \geq 2013)_t \\ &+ \beta_3 \mathbb{1}(g = G_{ct}^T) + \gamma_c + \epsilon_{gcdt}, \end{aligned} \tag{1}$$

Here, r_{gcdt} represents the retention rate in grade g of campus c in district d, year t, $\mathbb{1}(t \geq 2013)$ represents the year in which STAAR ratings first came into effect, and G_{ct}^T represents the final grade offered at school c in year t. γ_c represents school fixed effects.