Initial Results from the Third Grade Reading Guarantee Analysis [1]

Ani Ruhil, Ohio University
Lisa Neilson, Caroline Barto, Josh Hawley, Ohio State University
June 2023

Background

Ohio Excels has partnered with the Ohio Education Research Center at The Ohio State University to conduct a study of students who were retained by Ohio's Third Grade Reading Guarantee (TGRG). The Third Grade Reading Guarantee is based on changes to Ohio law in 2012, and since that time students were required to be on track to proficiency in reading during third grade to be promoted to the fourth grade. While the legislature authorized the State Board of Education to set the cut scores – the level at which a student must score or exceed to be promoted – it also expected the cut score to increase each year by law until it reaches proficient.

Critically, the legislation also outlined assessment plans for grades K-3 to ensure that students were monitored to see if their reading performance was on-track to achieving reading proficiency prior to the third grade. Intervention and progress monitoring are key components of these plans before the third grade. Plans are required for students that are not on track for reading proficiency, as determined by an assessment given at the beginning of each school year. For students subject to the reading retention guarantee, there are provisions for reading improvement and monitoring plans (RIMP). In 2013 the legislature made changes to the policy to ensure that students who did not pass the reading guarantee would be able to pass using an alternative or equivalent level of achievement. The 2013 policy change also exempted certain English learners (EL) and students with significant disabilities from Ohio's Third Grade Reading Guarantee (TGRG).

Practically speaking, the law has seen significant differences in the cut score and in the primary test used to make decisions about performance. In years 2014 and 2015 the third-grade reading (OAA) scaled score was used to make this decision. However, in 2016, 2017, 2018, and 2019 the OST third-grade ELA score was used to make the determination, while the reading sub score for the English language arts (ELA) assessment was an alternative measure used to promote students. In 2020, 2021, and 2022 neither test was used due to the pandemic emergency legislation.

In recent years, significant questions have been raised about the long-term effectiveness of retaining students. First, how many students were retained? Second, what were the demographics of the students retained? Third, does the policy yield the intended result, i.e., do retained students subsequently achieve proficiency because they were retained? And, finally, are there any unintended effects of retention on students who were retained?

These questions were informed by a review of prior literature on the impact of retention on students. System-wide retention has expanded significantly since 2002 when Florida enacted its policy of retaining third graders. By 2022, some 18 states had mandatory retention laws for students in third grade. The evidence for state-wide retention in early grades is, however, mixed. Some studies have identified small negative or positive effects on student outcomes. For example, a recent study on Indiana's retention policy by Hwang & Koedel (2022) concludes that "…*Indiana's retention policy has large positive short-*

[1] This study was funded by Ohio Excels to the Ohio State University. Data were provided by the Ohio Department of Education (ODE). We thank the funders, especially Lisa Gray, Kevin Duff, and Cassandra Palsgrove for support. We consulted with ODE to make sure that the data are used appropriately (Project: OLDA20220009). All errors are the responsibility of the research team.

and medium-term effects on same-grade student achievement in math and English Language Arts (Hwang & Koedel, 2022: 2)." Similar effects were found in a study of Florida by Schwerdt, West & Winters (2017). In this study the authors found that the positive impacts of retention in third grade have "…*substantial short-term gains in both math and reading achievement*" (Schwerdt, West & Winters (2017:155). Other studies have found significant losses in student academic achievement and negative social-emotional outcomes (Jimerson, 2001). While some of these differences in conclusions are likely driven by the methods of analyses used by the researchers, studies using a regression discontinuity model tend to find less negative, even some positive results, for retention on subsequent grade-level student performance (Hwang & Koedel, 2022).[2]

Data

This study uses data provided by the Ohio Department of Education to the Ohio Education Research Center (OERC). The data used include de-identified student level data with key information on student attributes (e.g., school district, individual gender/race, attendance, discipline) and test score information (e.g., Kindergarten readiness, ELA tests in third through eighth grade). Data on reading diagnostic tests was also provided for students from earlier grades. Data were provided to the OERC with a flag for students retained per the TGRG, so that the underlying data used for the study generally match the counts used by the State Department of Education for the number of students retained or not in a given year.[3]

Results

OERC conducted analyses of Education Management Information System (EMIS) student records to produce two sets of deliverables: descriptive statistics of the student population subject to the TGRG over time, and a study of the outcomes of students whose third-grade reading performance triggered the prescribed outcomes of the TGRG, as well as the performance outcomes for students whose third- grade reading performance was above the promotion cut score. This later analysis, when restricted to students who scored within a narrow range either side of the cut score, is labeled a Regression Discontinuity analysis.

Descriptive Findings: Using the data provided by ODE, from 2014-2019, a total of 20,895 third-grade students subject to the TGRG were retained per TGRG retention criteria. This is compared with 624,475 students who were not retained from 2014-2019 in their third-grade year because they met the TGRG passing threshold. There were an additional 39,654 students (2014-2019) who were NOT retained due to adequate performance on an alternative reading assessment, and another 1,315 students who were not retained because they met other promotion criteria of the TGRG policy.

[2] Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. Journal of Public Economics, 152, 154-169; Hwang, N & Koedel, C. (2022). Holding Back to Move Forward: The Effects of Retention in the Third Grade on Student Outcomes (Working Paper No. 22-688; Annenberg Institute at Brown University). https//doi.org/10.26300/mmxx-3e82; Jimerson, S. (2001). Meta-Analysis of School Retention Research: Implications for Practice in the 21st Century. School Psychology Review, 30:2, 420- 437

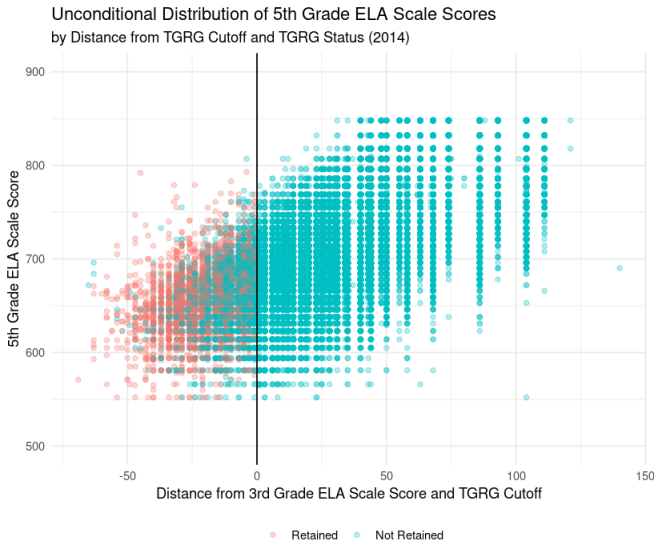
[3] Such differences as might exist are negligible and driven by differences in the data at varying points in time as they are updated when necessary.

The retained students were between 2.7% to 4.0% of all students subject to the retention policy. Numerically the largest group were retained in 2017 (4,590) and the smallest in 2016 (2,892).[4] Overall, some 55% of retained students were male (versus 50% of not retained students), and 91% were economically disadvantaged (versus 50% of not retained students). Of the 20,870 retained some 17% had a disability (versus 10% of not retained students). In terms of race and ethnic characteristics, the largest fraction (48%) of students retained were African American (versus 14.3% of not retained students), 34% were White, Non-Hispanic (versus 72% of not retained students), 11% were Hispanic (versus 6% of not retained students), and 7% were Multiracial or Other Races (versus 5% of not retained students).

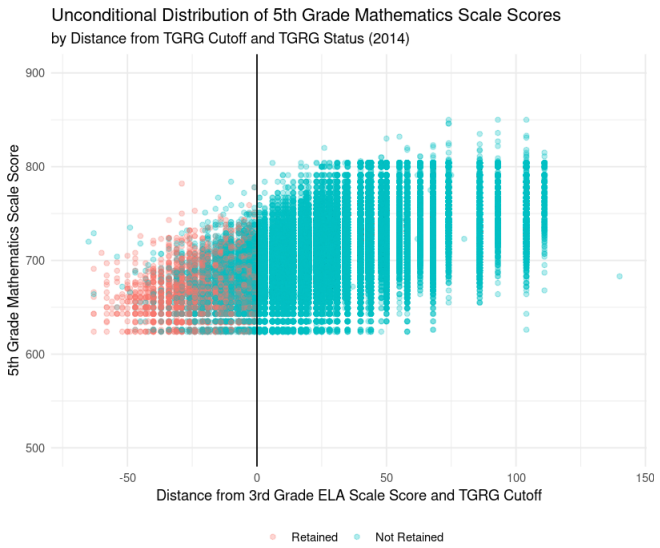
The retained students are almost all from the limited category of performance. Of the students retained, some 90% scored at the limited performance level in the qualifying assessment. Only 10% performed at the basic level. In comparison, only 7% of those not retained performed at the limited level and some 16% were at the basic level.

There are systemic differences in subsequent years in academic and performance between TGRG retained and not retained students in all years. For example, on achievement tests between fourth through the eighth grades, the retained students were less likely to be at or above proficient in math and reading across all years. Secondly, there are gaps in performance between retained and not retained students when comparing performance in subsequent grades by gender, race or ethnicity, and EL status.

It is important to remember that these raw differences in scores after the third-grade year are comparing **all** students who were not retained vs. **all** students who were retained. The figures below (Figure 1 and Figure 2) illustrate this well. In both cases, the reading and math scores for fifth grade retained and not retained students are displayed for **all** students from 2014. The students to the left of the cutoff are those with scores <u>below</u> the reading score required for retention in 2014, and the students to the right are those with scores <u>above</u> the cut score. Students who were promoted on time in third grade based on their reading scores display systematically higher scores on both the third and fifth grade outcome tests.



[4] The promotion score was higher in 2017 compared to 2018 and 2019.



Retake Results: Of the 20,895 students who were retained as a result of the Third Grade Reading Guarantee policy, 12,701 students retook the third-grade reading assessments in the following school year[5]. Within that group, 90% increased their assessment scale score in the subsequent year, and about half (53%) increased their reading performance. Twenty-one percent achieved reading proficiency (Table 1).

Table 1. Post-Retention Third-Grade Reading Performance Level

TGRG Year	Number of Students	Increased Score	Increased Performance Level	Achieved Proficiency
2014	2,363	89%	66%	41%
2015[6]	2,244	99%	21%	7%
2016	2,109	86%	49%	18%
2017	3,367	88%	58%	18%
2018	2,618	87%	63%	23%
Total	**12,701**	**90%**	**53%**	**21%**

Regression Discontinuity Results: The preceding analysis reflects the descriptive results. However, additional work by the OERC focused on understanding the post-retention performance of students who are close to the cutoff and therefore presumably more comparable. It is important to avoid concluding that there is a difference in the performance of retained versus not-retained students without zeroing in on the groups that displayed similar scores on the third-grade reading assessment used for the retention decision. A simple reading of the tables and figures in this memo leads one to an inappropriate conclusion because it compares all students, those who met or exceeded the proficiency score and those who were far from achieving proficiency.

The standard method used for studying the impacts of third-grade retention is called a regression discontinuity model. A regression discontinuity model requires a hard and fast rule about, in this case, a

[5] Students who may have been promoted mid-year were not separated from this group.
[6] SY 2014 and 2015 used a reading test for promotion, and SY 2016 and on used a more comprehensive ELA test that was measured on a different scale.

student's promotion to the fourth grade, and a cut score that both determines promotion and is applied consistently. It is also important to note that this methodology is recommended by the U.S. Department of Education as an appropriate tool when one cannot carry out a randomized controlled trial.

The following results (Table 2) serve as summary findings of the causal impact of retention in the third grade in 2014 or 2015 on math and reading scores in the fourth through the seventh grades. The model used in this analysis is a two-stage instrumental variable regression discontinuity analysis conducted by grade and subject, meaning that for each grade there are two separate models – one for reading and the other for math.

For more efficient analysis, we have pooled the model for the 2014 and 2015 TGRG cohorts when looking at fifth grade and sixth grade reading and mathematics scores. However, for the fourth-grade outcomes, we restricted the analysis to the 2015 TGRG cohort because the PARCC assessments coming into effect in 2015 were scaled differently than pre-PARCC assessments. We also restricted our seventh- grade analysis to the 2014 TGRG cohort to avoid including any assessments conducted in Spring 2020 or the following year. Note also that the bandwidth (the margin of scores either side of the TGRG cutoff score) used for each grade and subject varied across analyses and was chosen via recommended statistical techniques.

Table 2: IV 2SLS Estimates

	4th Grade		**5th Grade**		**6th Grade**		**7th Grade**	
	Reading	**Math**	**Reading**	**Math**	**Reading**	**Math**	**Reading**	**Math**
distance	1.28 (0.06)***	1.26 (0.10)***	1.35 (0.05)***	0.84 (0.05)***	1.03 (0.03)***	0.77 (0.03)***	0.85 (0.06)***	0.79 (0.05)**
Non Retention	-43.16 (6.99)***	-36.95 (10.18)***	-30.19 (4.33)***	-18.49 (3.66)***	-17.76 (2.98)***	-12.18 (2.73)***	-8.49 (4.03)**	-10.04 (3.53)***
95% CI of Retention	-56.67, -29.45	-56.89, -17.01	-38.67, -21.72	-25.66, -11.32	-23.62, -11.91,	-17.53, -6.83	-16.40, -0.59,	-16.96, -3.11
Bandwidth	+/- 15	+/- 14	+/- 16	+/- 14	+/- 17	+/- 18	+/- 15	+/- 17
Years included	2015	2015	2014-15	2014-15	2014-15	2014-15	2014	2014
N	24,798	21,048	46,150	41,517	52,086	53,199	20,152	23,796

** p <= 0.05; *** p <= 0.01

The results provide clear impacts of the causal effects of retention on subsequent grade and subject level performance. In all grades – fourth through seventh – and both subjects – reading and mathematics – there is a statistically significant difference in the performance of retained versus not retained students.[7]

Retention leads to, on average, retained students within a specific bandwidth around the cutoff scoring between 8 to 44 scale score points higher in fourth, fifth, sixth and seventh grade reading and mathematics assessments *than students who were not retained but were within the same bandwidth around the cutoff score*. The positive impacts of retention on performance in the fourth, fifth, and sixth grades in both subjects are stronger than in the seventh grade.

Interpreting the size of the positive impacts in a particular grade is no easy task. What seems clear, however, is that the size of the gap in the fourth grade is the largest, at 43.16 scale score points. The fourth-grade math results are 36.95 scale score points higher for the retained students near the cutoff than the not retained students near the cutoff. The mean scale score in 2015 and 2016 according to ODE data was around 710, and standard deviation was about 50 points for fourth-grade reading. This implies

[7] Note that the lower limit of the 95% confidence intervals for estimated impacts in 7th grade reading is very close to zero.

that a 40+ scale score point difference between retained and non-retained students is almost one standard deviation difference in mean scale score of retained versus not retained students. The mean scale score for math was also around 700 in 2015 and 2016 (the standard deviation for math was lower, approximately 36 points in fifth and sixth grades). Therefore, the difference in math scale score in fifth and sixth grades between those retained and not retained is approximately half a standard deviation.

Another way of examining differences in the size of the gap is to think about what would happen to an individual student if they had a certain scale score in fourth grade in terms of the performance level. Based on this logic if the 90% of students that performed at a limited level and were held back (almost all the students subject to the TGRG and retained had a score in the limited range), with a 40+ point increase in a scale score in fourth grade reading or math would raise the student to at the lowest level to a basic, and at the highest level to a proficient level. The limited scale scores fall between 545 and 671, and proficiency begins at 700 scale score points. If a student was at a basic level, a 40+ scale score increase would move them to high proficient or low accomplished range. As the scale score differences get smaller, 17 points in sixth grade reading for example, the impact of retention three years earlier are smaller, although they still could move a student from basic to proficient or from the low end of limited to a higher level.

Appendix B provides detailed figures with information on the distribution of reading and mathematics scores for retained and not retained groups.

Conclusions

The memo summarizes initial results from a study of the effects of retention on Ohio third graders. Using data from the Ohio Department of Education provided for this study to the Ohio Education Research Center, we document that the vast majority of Ohio students from 2012-2019 who took the third-grade reading assessments and were subject to potential retention were promoted. Some 20,895 students were retained in the 2014-2019 cohorts, in comparison to 624,475 students who were not retained. A further 39,654 students were promoted using an alternative reading assessment.

There are systemic differences in subsequent years in academic performance between TGRG retained and not retained groups in all years. The academic performance in fourth through eighth grade on subsequent reading and mathematics tests continues to show significant differences between those that were retained and the much larger group of students that were not retained. These differences in academic performance hold across all grades we tested, regardless of demographic categories.
However, these systemic differences are present when comparing **all students** vs. the small subgroup that were retained. The all students not retained group contains those scoring at all proficiency levels, from accelerated to not passing. Therefore, the results require a research approach that will allow us to focus on comparing similar groups.

The regression discontinuity model we carried out uses data from the 2014 and 2015 TGRG cohorts, and examines individuals retained and not retained in a very narrow bandwidth around the score needed for retention in either of those years. Using this analytical framework, we can show a causal estimate of the impact of retention on student academic performance in the fourth through seventh grade (data from eighth grade is not yet available for the 2014 cohort). In these models we observe between a half and a full standard deviation for students retained in subsequent reading and math performance in fourth through seventh grades.

Appendix A

All data utilized in this analysis were provided by the Ohio Department of Education (ODE). Other information was also gathered via online documents or then provided by ODE, in particular documents summarizing the technical details of state assessments, cut scores for proficiency levels, and third grade reading guarantee promotion scores.

Table A1: Tests, Promotion Score and Proficiency Score (2014-2019)

	Test Type	**Promotion Score**	**Proficiency Score**
2014	OAA	392	400
2015	OAA	394	400
2016	OST	42 Reading sub-score	50 Reading sub-score or 700 ELA scaled score
2017	OST	44 Reading sub-score	700 ELA scaled score
2018	OST	44 Reading sub-score or 672 ELA scaled score	700 ELA scaled score
2019	OST	45 Reading sub-score or 677 ELA scaled score	700 ELA scaled score

The OERC analytics team constructed a unified dataset that combined student-level demographics with TGRG indicators and scaled scores on OAA/Next Generation assessments[8] by grade and subject. The resulting dataset contained records for 753,998 students.

To define students as TGRG Retained or TGRG Not Retained, we used ODE-provided indicators for whether a student had been subject to the TGRG retention and retained, or subject to the TGRG retention but not retained (Table 21). The analysis excludes students who were not subject to the TGRG and students who were retained for non-TGRG reasons. Also, cases with inconsistent TGRG information or incomplete proficiency information are excluded from the analysis.

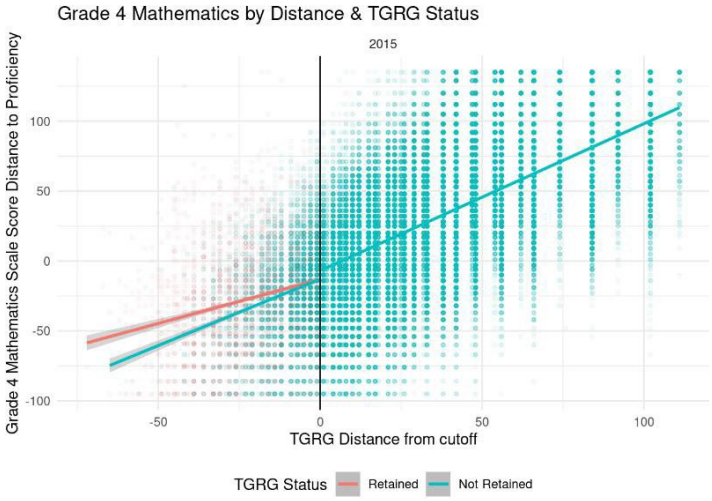
Table A2. Analytic Sample

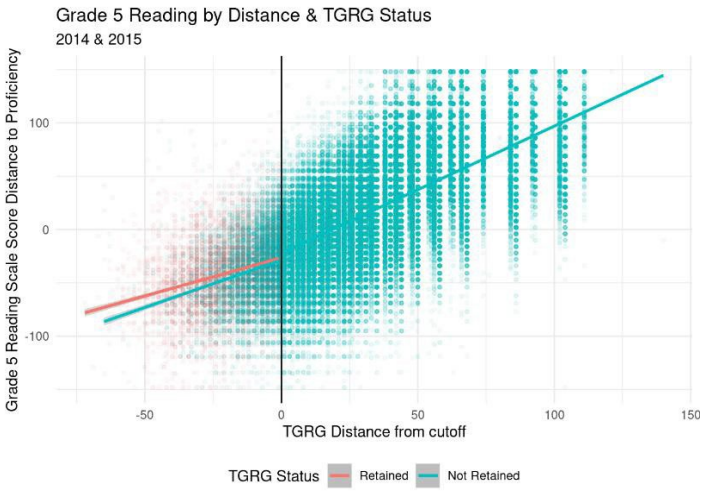
Description	n
Student was retained at the end of the previous school year due to the TGRG and is still retained.	19,515
Student was retained at the end of the previous school year due to the TGRG but is not enrolled this school year.	1,370
No retention reason reported in EMIS	10
Total TGRG Retained9	20,895
Student was not retained at the end of the previous school year.	624,475

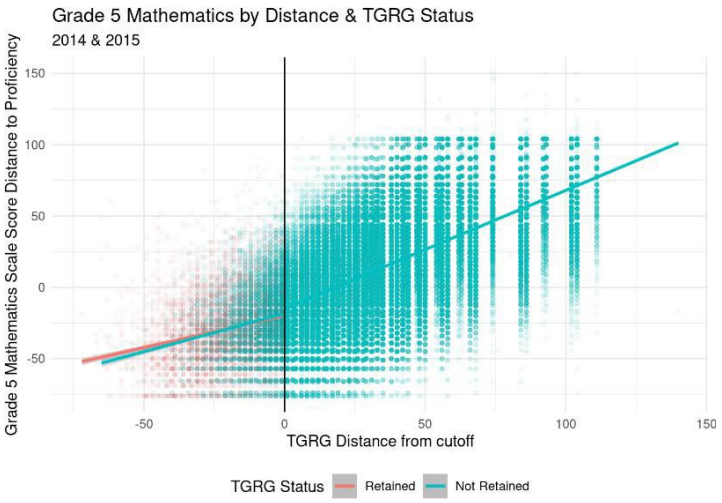
[8] ODE switched away from the OAA for Reading in the 2015-2016 school year. At the same time, Ohio's definition of "proficient" changed to a more rigorous standard a broader set of skills in English Language Arts.

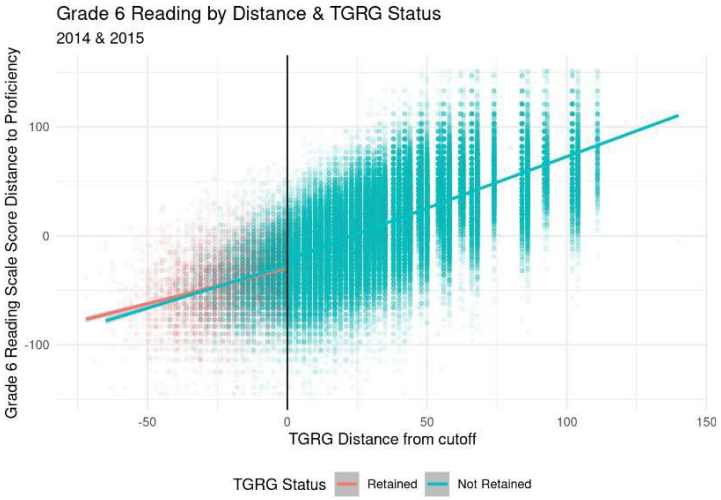
[9] Includes 25 errant records that are omitted from the final analytic sample.

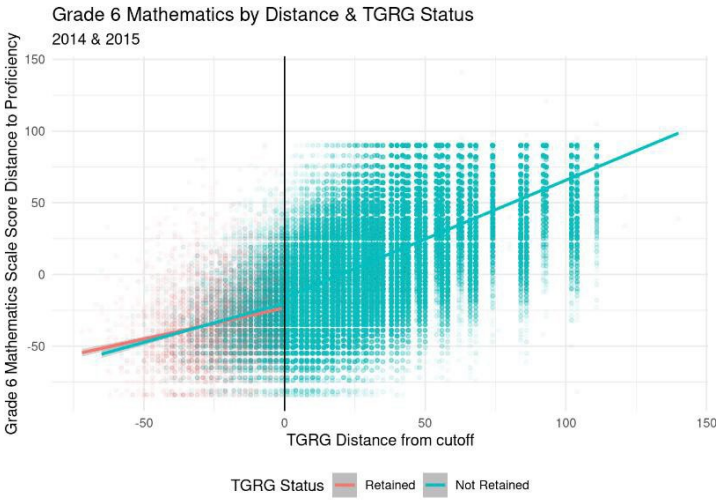
Appendix B: Reading and Mathematics in fourth-seventh grades (2014-2015 cohorts, pooled estimates)

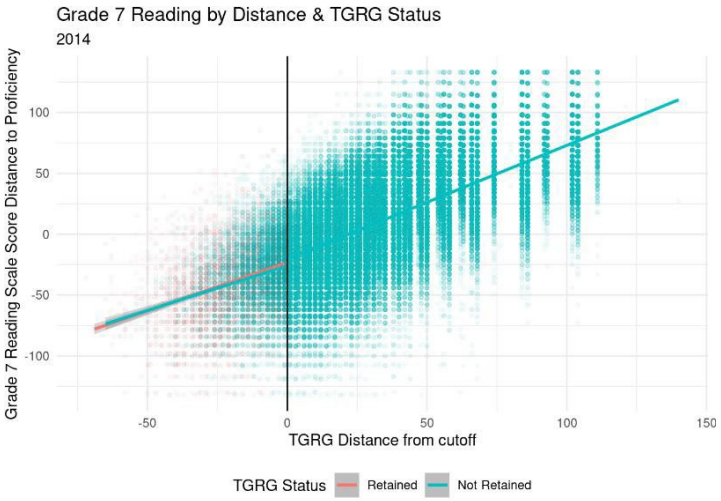


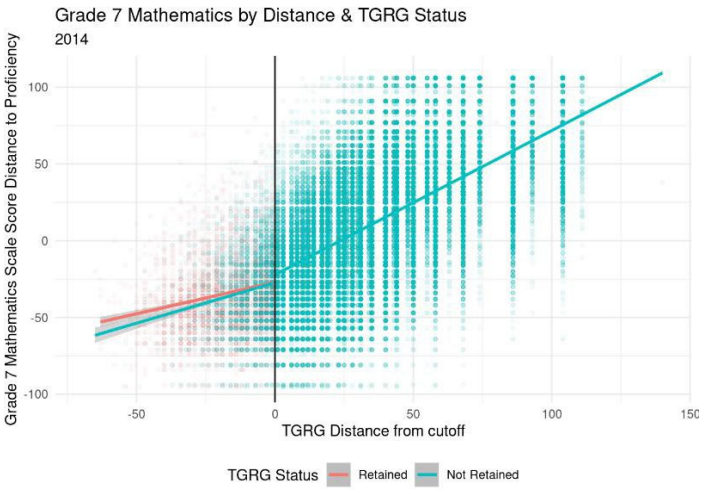



Grade 5 Mathematics by Distance & TGRG Status
2014 & 2015
Grade 5 Mathematics Scale Score Distance to Proficiency
150
100
50
0
-50
-50
0
50
100
150
TGRG Distance from cutoff
TGRG Status
Retained
Not Retained


Grade 6 Reading by Distance & TGRG Status
2014 & 2015
Grade 6 Reading Scale Score Distance to Proficiency
100
0
-100
-50
0
50
100
150
TGRG Distance from cutoff
TGRG Status
Retained
Not Retained


Grade 6 Mathematics by Distance & TGRG Status
2014 & 2015
Grade 6 Mathematics Scale Score Distance to Proficiency
150
100
50
0
-50
-50
0
50
100
150
TGRG Distance from cutoff
TGRG Status
Retained
Not Retained


Grade 7 Reading by Distance & TGRG Status
2014
Grade 7 Reading Scale Score Distance to Proficiency
100
50
0
-50
-100
-50
0
50
100
150
TGRG Distance from cutoff
TGRG Status
Retained
Not Retained


Grade 7 Mathematics by Distance & TGRG Status
2014
Grade 7 Mathematics Scale Score Distance to Proficiency
100
50
0
-50
-100
-50
0
50
100
150
TGRG Distance from cutoff
TGRG Status
Retained
Not Retained