



Is Read to Achieve Making the Grade?

An Assessment of North Carolina's Elementary Reading Proficiency Initiative

October 2018

Sara Weiss and D.T. Stallings
The William and Ida Friday Institute for Educational Innovation

Stephen Porter
College of Education

North Carolina State University

NC STATE

College of Education

Acknowledgments

This research project has been funded by the Institute for Education Sciences under the Low-Cost, Short-Duration Evaluation of Education Interventions program, grant number R305L160017. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Table of Contents

Summary	4
Introduction	6
Background: Overview of Read to Achieve	7
Goal	7
Determination of Proficiency	8
Interventions and Supports	8
Retention and Promotion	9
Purpose of the Report	9
Partners	9
Is There Evidence the Policy Could Work?	10
Our Approach to the Analysis	11
Data	11
Analysis	11
Outcomes of Interest	11
How We Estimate Impact	12
Student Outcomes One and Two Years after Initial Read to Achieve Identification	16
What is the causal effect of the Read to Achieve program on student reading performance one year and two years later?	16
What is the causal effect of being retained in 3 rd grade on student reading performance one year later?	18
How do short- and longer-term effects vary by student sub-groups (e.g., gender, race/ethnicity, and economic disadvantage, means of demonstrating proficiency)?	18
Does participation in a reading camp make a difference?	19
Policy Implications: Why Are There No Positive Results?	20
Limitations of the Current Study	20
Potential Policy Challenges	21
Lack of Support for Pre-3 rd Grade Intervention	21
Broad Definition of Reading Proficiency	21
Assumptions about Availability of Qualified Teachers	21
The Nature of RtA Retention	22
Potential Implementation Issues	22
Differences in Local Camp Structure	22

Differences in Local Capacity	23
Variability in Interventions Ahead of the 4 th Grade Year	23
Variations in the Retained Student Experience	24
Promotion to Grade 5	25
Moving Forward: What the State Can Do Next	26
Improve Implementation Fidelity	26
Identify and Scale Up Local Successes	26
Transition from a Social Promotion Mindset to a Literacy Development Mindset	27
References	28
Appendix A. Intervention Flowcharts	31
Appendix B. Review of Recent Research on the Effectiveness of Read to Achieve Components	33
Appendix C. Outputs from Statistical Models	35

Summary

Context

In 2012, in an effort to increase elementary reading achievement in North Carolina, and to end a *de facto* policy of “social promotion” that places more emphasis on age than on demonstrated proficiency, the North Carolina General Assembly passed legislation that required the North Carolina Department of Public Instruction to develop and implement a program to support on-grade reading mastery for all 3rd grade students. The initiative is commonly referred to as Read to Achieve (RtA).

The RtA policy provides multiple supports for students who do not demonstrate reading proficiency by the end of 3rd grade, including an optional reading camp between the 3rd and 4th grade years. For students who do not become proficient by the end of the summer, supports include supplemental tutoring and enhanced reading instruction during the next school year. Implementation occurs at the school district level but is funded primarily by the state.

After five full years of implementation,¹ has the investment been worth it? The history of 3rd and 4th grade End-of-Grade (EOG) reading scores has not been promising, with test scores either remaining relatively flat (4th grade) or even slightly declining (3rd grade) since the start of the program; however, global measures like those may hide important gains for the 3rd and 4th graders most directly impacted by the policy.

To begin to uncover the academic impacts of RtA, this report presents analyses of 4th and 5th grade reading test scores for the state’s traditional public school students who first experienced the RtA initiative as 3rd grade students during the 2013-14 and 2014-15 school years.

Student Outcomes One and Two Years after Initial Read to Achieve Identification

Causal Effect of Initial Identification on Student Reading Performance

Our initial analyses include all students *originally impacted by the initiative*—that is, all students who do not demonstrate proficiency in reading on their first 3rd grade EOG. These students are eligible for RtA services between 3rd and 4th grade, but only some end up being retained (those who do not demonstrate proficiency even after exposure to the supports and before the start of 4th grade).

The improvements in EOG scores for these initially identified students both one year and two years after identification are not statistically different from the improvements in EOG scores of comparison students who receive no RtA services. Entering the RtA process after completion of the 3rd grade EOG and before enrollment in 4th grade does not appear to have had an impact on the first two cohorts of RtA students.

¹ 2013-14 was the first year of implementation.

Causal Effect of Retention on Student Reading Performance

Our second set of analyses include only students who are identified as retained at the start of what would have been their 4th grade year—that is, students who do not demonstrate proficiency in reading after interventions between the 3rd grade EOG and the start of the next school year.

Results of these analyses also suggest that there are no significantly different outcomes for this group, one or two years later, compared to outcomes for students who just missed identification and received no additional services.

Effects for Student Sub-groups

Our analyses also considered the impact of the policy on sub-groups of students—for example, students from lower-income families, male and female students, and students from different ethnic groups. Even looking at results for these sub-groups, however—whether just initially identified or eventually retained—there still does not appear to be any effect.

Reading Camps

One key component of the policy—participating in reading camp between the 3rd and 4th grade years—is not mandatory. As a result, each year, a large proportion of students who are eligible for RtA services are not exposed to one of its major interventions. We compared outcomes for students who attended reading camps to outcomes for students who were eligible for reading camps but did not attend. As was true in all of the other analyses, participation in a reading camp does not appear to make a difference in subsequent test scores.

Moving Forward: What the State Can Do Next

One reason for the general lack of overall progress may be the significant gaps between the RtA policy (such as the policy's broad definition of reading proficiency and the assumptions it makes about the statewide availability of high-quality reading teachers) and several aspects of on-the-ground implementation realities (such as differences across school districts in program offerings and staff capacity, or variations in the services offered to retained students). To strengthen the program, the state should: 1) consider providing the financial and human capacity supports necessary to improve implementation fidelity statewide; 2) identify and scale up local-level implementations with strong evidence of success; and, ultimately, 3) consider transitioning from a 3rd grade social promotion mindset to a literacy development mindset that spans all education settings leading up to and including 3rd grade.

Introduction

Retaining elementary students is growing in popularity nationwide as a policy approach to ensure that students are prepared for upper elementary school and beyond. In 2012, in an effort to increase elementary reading achievement in North Carolina, and to end a *de facto* policy of “social promotion” that places more emphasis on age than on demonstrated proficiency, the North Carolina General Assembly passed legislation² that required the North Carolina Department of Public Instruction (NCDPI) to develop and implement a program to support on-grade reading mastery for all 3rd grade students. The initiative is commonly referred to as Read to Achieve (RtA).

The RtA policy provides multiple supports for students who do not demonstrate reading proficiency by the end of 3rd grade, including an optional reading camp between the 3rd and 4th grade years. For students who do not become proficient by the end of the summer, supports include supplemental tutoring and enhanced reading instruction during the next school year. Implementation occurs at the local education agency (LEA)³ level but is funded primarily by the state; in 2015, the General Assembly even expanded its financial commitment by \$20 million a year.

While similar to policies in other states (such as Florida), RtA establishes a higher threshold for proficiency than most. At the end of the first school year after implementation of RtA (2013-14), almost 46,000 3rd grade students (about 40%) initially were identified as not demonstrating reading proficiency, and nearly one-third of those still did not demonstrate proficiency by the start of the 2014-15 school year.

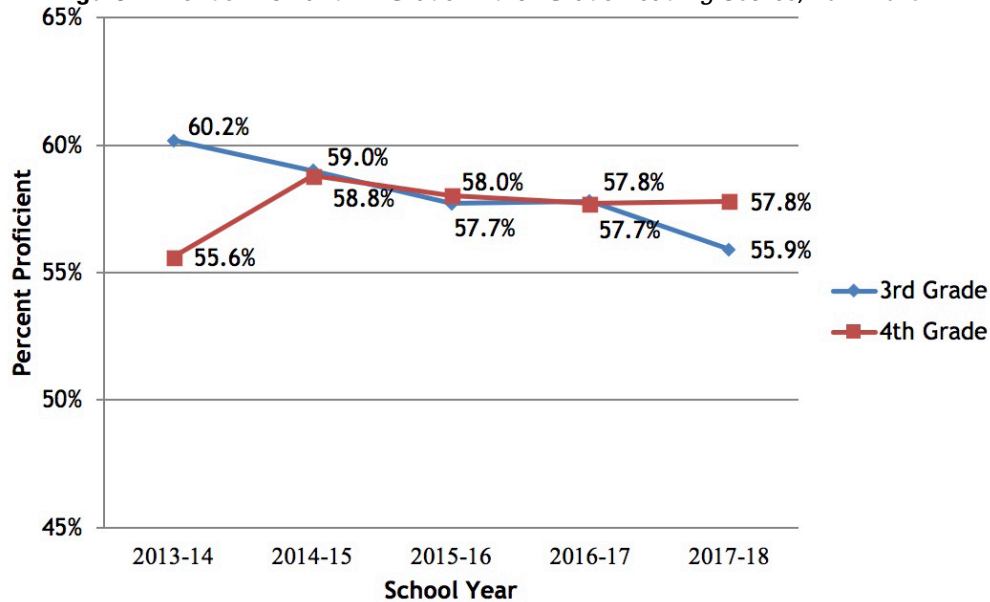
With so many students and schools potentially impacted every school year, and with such a significant investment in terms of both time and money, RtA arguably is one of the most influential education policies currently in place in North Carolina. But after five full years of implementation, has the investment been worth it? The history of 3rd and 4th grade End-of-Grade (EOG) reading scores has not been particularly promising (Figure 1, following page), with test scores either remaining relatively flat (4th grade) or even slightly declining (3rd grade) since the start of the program; however, global measures like those may hide important gains for the 3rd and 4th graders most directly impacted by the policy.

To begin to uncover the academic impacts of RtA, this report presents analyses of 4th and 5th grade reading test scores for the state’s traditional public school students who first experienced the RtA initiative as 3rd grade students during the 2013-14 and 2014-15 school years.

² NC Session Law 2012-142, House Bill 950, <http://www.ncleg.net/Sessions/2011/Bills/House/PDF/H950v7.pdf>

³ LEA is North Carolina’s term for school district.

Figure 1. Trends in 3rd and 4th Grade End-of-Grade Reading Scores, 2014-2018



Background: Overview of Read to Achieve

Goal

The goal of RtA is to ensure that every North Carolina public school student reads at least at grade level by the end of 3rd grade. The complex policy has three main components:

- 1) Determining whether students are proficient in reading by the end of their 3rd grade year;
- 2) Providing multiple interventions (including local reading camps) in the spring and summer for students who are not proficient; and
- 3) Retaining or promoting students before the beginning of what would be their 4th grade year, if they are not able to show proficiency before the beginning of that school year.

The various pathways students follow to proficiency under the policy are detailed in the flowcharts in **Appendix A**.

Determination of Proficiency

In the spring of their 3rd grade year, all students take an EOG test to measure their reading proficiency. If a student does not demonstrate proficiency on the EOG, the student has the option to demonstrate proficiency via additional assessments, including an EOG re-test, an alternate assessment known colloquially as the “RtA test,”⁴ locally-developed assessments, or a reading portfolio that demonstrates student proficiency via reading passages and questions that cover 3rd grade reading standards. In addition, a student’s Beginning-of-Grade reading pre-test score also can count as a measure of grade-level proficiency, even if her or his EOG score later that year is below proficient. Students without any “good cause exemptions”⁵ who do not score at or above the proficient level on any of these tests are designated for retention, but retention is not finalized until just before the beginning of the following school year.

Interventions and Supports

Each of the state’s 115 traditional LEAs is required to provide eligible students with access to a reading camp between 3rd and 4th grade.⁶ The camp must offer at least 72 hours of reading instruction by teachers with demonstrated positive student outcomes in reading proficiency.⁷ If the student’s family declines enrollment in the camp, the student is retained in 3rd grade automatically for the following year.⁸ If the student attends and completes the reading camp, the student can demonstrate proficiency via one of the measures described above. Students who demonstrate proficiency at the end of reading camp are promoted to 4th grade.

Retention and Promotion

Students who attend reading camp but do not demonstrate proficiency by the end of camp are identified as retained and are placed in one of three special school settings at the start of the next academic year: a 3rd grade accelerated reading class with 90 additional minutes of reading instruction per day; a hybrid 3rd grade/4th grade transition class with 90 additional minutes of reading instruction per day, but in a 4th grade setting; or a 4th grade accelerated reading class—a regular 4th grade placement but with 90 minutes of pull-out reading instruction per day.⁹

⁴ An assessment developed by NCDPI that is similar in rigor to the EOG. The main difference between the RtA test and the standard EOG is that the EOG is comprised of long reading passages with questions at the end, while the RtA test breaks the longer passages into smaller units with questions interspersed between the passages.

⁵ Students who are designated as having limited English proficiency or a reading-related learning disability, or who have been previously retained for reading issues.

⁶ Reading camps are financially supported by legislatively-allocated funds. Charter schools are not required to provide summer reading camps, though some do and receive funds to do so.

⁷ In North Carolina, demonstration of positive student reading outcomes is determined by a statewide value-added assessment known as the Education Value Added Assessment System, or EVAAS. EVAAS scores for upper elementary teachers are based primarily on student performance on the state EOG.

⁸ Parents also can opt to enroll their students in a private version of the reading camp and then test their children for proficiency alongside children who attend the district’s reading camp.

⁹ Students in the 4th grade special placement are still identified as retained until they demonstrate on-grade reading proficiency.

During the retention year, starting in November, RtA students can apply for formal promotion to 4th grade by demonstrating proficiency via the pathways outlined above. Promotion to 5th grade depends only in part on performance on the 4th grade reading EOG. If students are proficient on the 4th grade EOG, they are eligible to be promoted to 5th grade, regardless of their grade level at the time of the test; however, if students are not proficient on the 4th grade EOG, school principals determine whether students are retained.

Purpose of the Report

Clearly, RtA is a wide-reaching policy with a number of moving parts; getting a better understanding of the impact of each of those components is not straightforward. The purpose of this report is to introduce one method for doing so at the state level and to begin to unpack the results of assessments of the impact of the policy on the first two cohorts of 3rd grade students exposed to it. To do so, we developed three overarching questions to guide our work:

1. *What is the causal effect of the Read to Achieve program on subsequent student reading performance one year and two years later?* In other words, what happens to the reading scores of students who are initially identified for possible retention as a result of their 3rd grade reading EOG scores?
2. *What is the causal effect of being retained in 3rd grade on student reading performance one year later?* In other words, what happens to the reading scores of the sub-set of students who, by the start of what would have been their 4th grade year, actually *are* retained?
3. *How do short- and longer-term effects vary by student sub-groups?* That is, are outcomes different for students based on their gender, ethnicity, economic status, or even based on the way(s) in which they demonstrate proficiency after initial identification?

Partners

Understanding the nuances of a complex program like RtA is not achieved simply by reading legislation and analyzing data; it requires an ongoing and open exchange among policymakers, implementers, and researchers. The partners involved in our study represented the following organizations:

- The North Carolina Department of Public Instruction (especially the Division of K-3 Literacy);
- The Research and Evaluation Team at NCSU's Friday Institute for Educational Innovation; and
- The Department of Educational Leadership, Policy, and Human Development at North Carolina State University's College of Education.

Is There Evidence the Policy Could Work?

RtA represents a major shift in the state’s approach to early-grade retention in five ways. First, the policy includes a credible threat of retention, followed by a measurable increase in the actual retention rate of 3rd grade students, from under 2% annually—most of which were for non-reading reasons, such as attendance—to more than six times that. Third, the policy introduces the elementary school equivalent of a summer school in the form of reading camps, which focus on intensive reading instruction for a student group that otherwise would have no formal, state-provided learning opportunity between 3rd and 4th grade. Fourth, schools theoretically provide higher-quality classroom instruction in reading by placing retained students with teachers who have strong track records in reading instruction. Finally, retained students are exposed to more hours of reading instruction than are their 4th grade peers. Table 1 provides a quick summary of evidence from past research on the potential effectiveness of each of these components; **Appendix B** includes a more detailed review.

Table 1. Findings from Recent Research on the Potential Impact of RtA Components

Component	Findings	Sample References
<i>Threat of Retention</i>	Unclear; some evidence that older students work harder, but without positive academic results	Range et al., 2012; Roderick & Engel, 2001; Bandura, 2001; Roderick et al., 2002
<i>Early-Grade Retention</i>	Higher-quality studies suggest evidence of short-term positive effects; some evidence of longer-term carry-over	Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005; Greene & Winters, 2007; Mariano & Martorell, 2012; Schwerdt & West, 2013; Winters & Greene, 2012
<i>Reading Camps</i>	Evidence of positive gains, tempered by the quality of the program	Jacob & Lefgren, 2004; Mariano & Martorel, 2012; Cooper et al., 2000; Kim & Quinn, 2013
<i>Teacher Quality</i>	As defined by value-added metrics, teacher quality matters; other measures of quality are less predictive of impact	Chetty et al., 2011; Rivkin et al., 2005; Rockoff, 2004; Goldhaber & Brewer, 2000; Boyd et al., 2008; Dee, 2007; Ehrenberg & Brewer, 1995; Goldhaber & Hansen, 2009; Goldhaber & Hansen, 2013; Hanushek et al., 2005
<i>Additional Reading Instruction</i>	Unclear, but promising if sustained across grades	Gettinger, 1995; Marzano et al., 2000; Harn et al., 2008; McMaster et al., 2005; Vadasy et al., 2002; Vellutino et al., 1996; Rowan & Correnti, 2009

In sum, recent research has identified some potentially positive impacts on student achievement of many of RtA’s key components, when each component is studied in isolation, as well as some less-certain impacts. The RtA policy, if implemented with fidelity, has the potential to impact student reading proficiency through the collective impact of all five of these components.

Our Approach to the Analysis

This section provides a brief, non-technical overview of the data we collected and the analyses we used. We will release a working paper later this fall with more complete technical coverage of our data sources and methods.

Data

We used data from three sources:

1. *Public School Demographic and Testing Data.* Our two student cohorts of interest are school-year 2013-14 and 2014-15 3rd graders—the students who experienced the first two years of RtA and who have completed several years of schooling after that point. Student-level data were drawn from the 2013-14 through 2016-17 school years. In addition to test scores, we also used several demographic indicators such as gender, race/ethnicity, whether the student was a participant in the free or reduced-price lunch program, and whether the student was absent more than 10 days during the school year.
2. *Reading Camp Details Survey.* In 2017, we surveyed a Read to Achieve contact person in each LEA to learn more about the different ways in which reading camps were structured. We received partial or complete responses from 110 of 115 LEAs.¹⁰
3. *Elementary School Placement Survey.* In 2017, we also surveyed every elementary school principal to get a better understanding of the diversity of placement options in their schools for students who were required to enroll in one of the RtA classrooms for what would have been their 4th grade year. We received responses from nearly 400 elementary schools (about 30%).

Analysis

Outcomes of Interest

Our primary outcome of interest in most of our analyses is **students' reading EOG scores** one year and two years after they take the 3rd grade reading EOG test for the first time. The reading EOG test content varies slightly across 3rd, 4th, and 5th grades, but in all three cases is comprised of about 35% reading for literature, 45% reading for information, and 20% language. Unlike some state standardized exams, the reading EOG is *vertically scaled*, which means that scores from one year are on the same scale as scores from another year and can be compared to each other (that is, a score of 420 on a 3rd grade test indicates the same reading ability as a score of 420 on a 4th grade test). This type of test is particularly useful for this analysis, as it allows us to compare test scores for all students in the same 3rd grade cohort, regardless of their grade designations in the years following 3rd grade (Roderick & Nagaoka, 2005).

¹⁰ The five LEAs that did not respond to any of the survey questions were Anson County, Cherokee County, Lexington City, Franklin County, and Halifax County.

A secondary outcome of interest is **3rd grade retention**, which we define as beginning the school year after 3rd grade without having demonstrated reading proficiency on the 3rd grade EOG or any of the alternate examinations. Thus, even a student who is promoted to 4th grade at or after the initial November reconsideration date in the school year following 3rd grade is counted as “retained” in our analyses, as is a student who has not demonstrated reading proficiency but still is placed in a 4th grade classroom at the start of the school year.¹¹ Our analyses do not directly estimate the impact of RtA on retention, but we are interested in how RtA impacts various subgroups of students—most especially those who are officially retained at the start of what would have been their 4th grade year.

How We Estimate Impact

The most important thing to keep in mind about our analyses is that our goal is to determine *causal*, not correlational, effects—that is, effects that we can directly attribute to the policy itself, not effects that happen to be numerically correlated with certain aspects of that policy. Because it is usually very difficult to achieve truly experimental conditions in school settings, most education researchers typically have to fall back on correlational analyses; however, due to the way RtA is implemented, we had a rare opportunity to conduct many analyses that are nearly experimental in nature, which allows us to arrive at more directly causal conclusions.

The key feature that allowed us to do this is that all students in 3rd grade (without a good cause exemption) are equally eligible for the program the day before they take their 3rd grade EOG. In other words, the entire population of students has the potential to be identified for RtA services, and that identification is based on one variable only: the 3rd grade EOG score.

Because the only factor that makes a student eligible for RtA is a single test score, all of the other characteristics of students with scores *just above or below the eligibility score* should be present in essentially the same proportions. The difference between scoring, say, a 438 or a 439 on the 3rd grade EOG is a matter of answering one question correctly or incorrectly, and chances are strong that no other characteristic is systematically or consistently different, on average, for the entire group of students who score 438 and for the entire group of students who score 439. In other words, the two groups are as close as we can come to a truly random experimental and control group as is possible without actually running a formal experiment. The only collective difference for these students is that the students who scored just below the passing threshold for readings are exposed to the host of RtA interventions, while those who scored just above the passing threshold are not.

This situation allows us to conduct *regression discontinuity* analyses, with the only difference between the two groups at the RtA cut-line being the single change in conditions (RtA eligibility status) immediately after the test is taken. Since all else about the two groups is, on average, the same, we can make strong conclusions about whether RtA eligibility *on its own* directly contributed to our outcomes of interest. What we are looking for is evidence that students just on either side of the EOG cut-point perform differently from each other at the end of the first and second years after 3rd grade. The premise is that they have essentially the same reading ability, but only some of them are impacted randomly (i.e., because of a random response to a single test question) by RtA.

¹¹ **Appendix A** illustrates all of the possible placement options for retained students.

As described before, students can be exposed to a number of different combinations of interventions immediately after schools learn their 3rd grade EOG scores. We can identify most of these intervention combinations for each student, but because each combination essentially represents a different “experiment,” the strongest conclusions we can make are about the impact of being initially eligible for RtA services (after the first test but before any interventions). This analysis—called a *sharp* regression discontinuity—only considers the *impact of the presence of the entire policy* (beginning with the first possibility of retention), not the impact of specific interventions or combinations of interventions that happen after the first test (e.g., re-tests, summer camp, additional instruction, eventual retention, etc.). That is, students who initially score below the cutoff are then exposed to a multifaceted treatment called “RtA,” which may or may not include retention as a component.

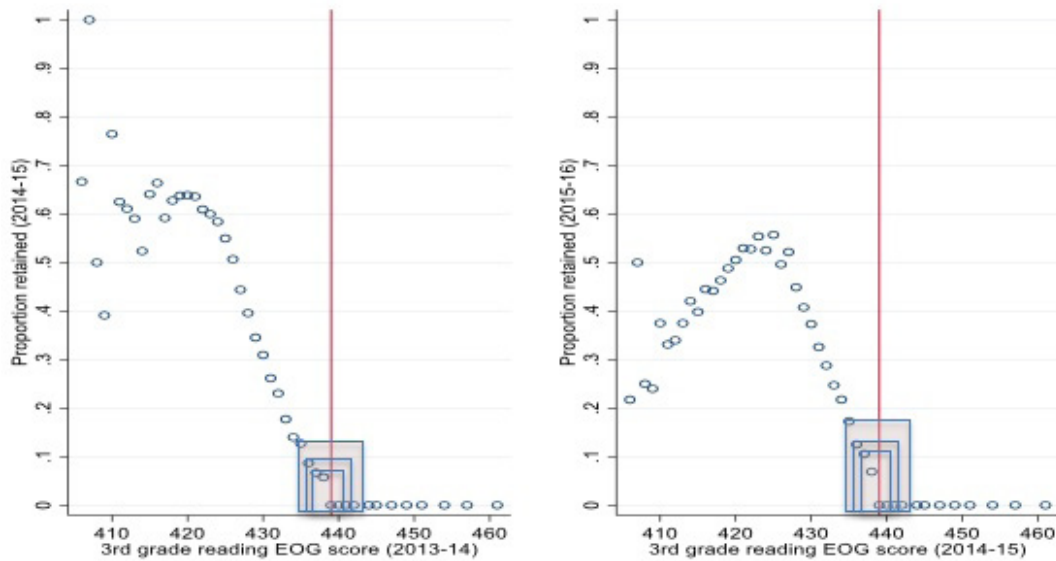
When we analyze outcomes for the subgroup of students who, after initial identification and any combination of interventions, are retained entering what would have been their 4th grade year, we are conducting a *fuzzy* regression discontinuity, since the post-EOG test experiences and conditions that ended in their eventual retention are different, even though all were originally identified for the same reason.

Figure 2 (page 15) helps to demonstrate the differences between the sharp and the fuzzy regression discontinuity analyses. The vertical red line represents the eligibility score for RtA; all students with scores to the left of that line are *eligible* for RtA, and all students to the right of that line are not. The small blue circles represent the proportion of students with each 3rd grade EOG score who eventually are *retained* by the start of the following school year. Note that in very few cases are all students with a certain score below the cut-line retained—primarily because many of them go on to demonstrate proficiency over the months following their first EOG via the alternative methods described earlier. The *policy* affects all of the students to the left of the line, but actual *retention* affects only some of the students to the left of the line; in other words, when we analyze the impact of retention, we lose some of the experimental strength we gained from the single-factor eligibility condition.

The three shaded boxes represent the students who are included in our analyses. In the smallest box are the students whose 3rd grade EOG scores are between two points above and two points below the cut-line; in the largest box are students whose EOG scores are between four points above and four points below the cut-line. We conducted our analyses on different sizes of groups to ensure that we are studying outcomes for the groups most likely to be exactly alike, other than their score (the group within two points of the cut-line) as well as outcomes for a more representative group of impacted students, even though their average group characteristics might be a little bit different (the groups within three and four points of the cut-line).¹² Focusing on this portion of the entire group of 3rd grade students does not affect the strength of the analyses because it includes about 30,000 students each year, but it does introduce one limitation to what can be concluded based on our analyses, which we discuss in greater detail in the **Policy Implications** section later in this report.

¹² The primary analysis results presented in this report are for the largest group— ± 4 points—unless otherwise noted. Results for the more narrowly-defined but less-representative analyses (± 3 points and ± 2 points) were the same and will be included in the forthcoming technical report.

Figure 2. Relationship between Initial 3rd Grade Reading EOG Scores and Eventual Retention



Note: Likelihood of retention initially increases as a student's initial score decreases, but the likelihood begins to decline again as the number of students with good-cause exemptions increases.

Student Outcomes One and Two Years after Initial Read to Achieve Identification

All of the outputs from our statistical analyses are included in **Appendix C**, along with explanations of how to read and interpret them. We have included graphic examples of some of those outputs in this section to illustrate what our analyses revealed.

What is the causal effect of the Read to Achieve program on student reading performance one year and two years later?

First, on the surface, and as mentioned in the **Introduction** (Figure 1), there do not appear to be any noticeable test score gains for students impacted by the program, whether we look at Cohort 1 (2013-14 3rd graders) or Cohort 2 (2014-15 3rd graders), one year out or two years out. The figures below (Figures 3 and 4) illustrate this by matching each possible 3rd grade EOG score with the average 4th and 5th grade EOG scores for students who earned each 3rd grade score. Note that, while scores for each group of 3rd graders are higher in the years that follow (indicating reading growth), the increase in scores across years for any group is about the same as the increase for all groups—in other words, most students who initially scored below the cut-line are not outgaining students who initially scored above the cut-line. As an example, students who scored 430 in 3rd grade in 2013-14 scored, on average, about 440 in 5th grade in 2015-16 (green dashes)—a ten-point increase; students who scored 440 in 3rd grade in 2013-14 scored, on average, about 450 in 5th grade in 2015-16 (red dashes)—also a ten-point increase.

Figure 3. 2013-14 and 2014-15 Cohorts, One Year Out

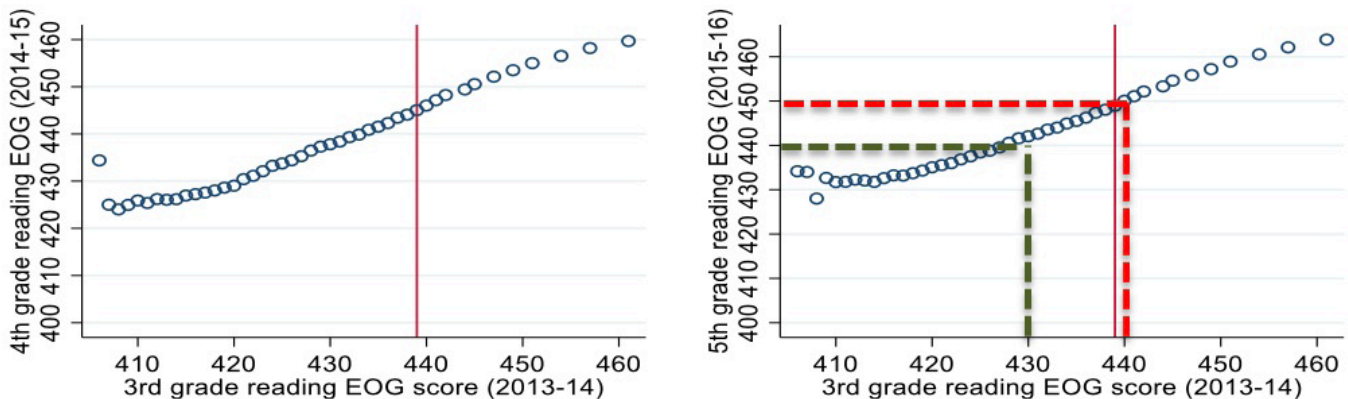
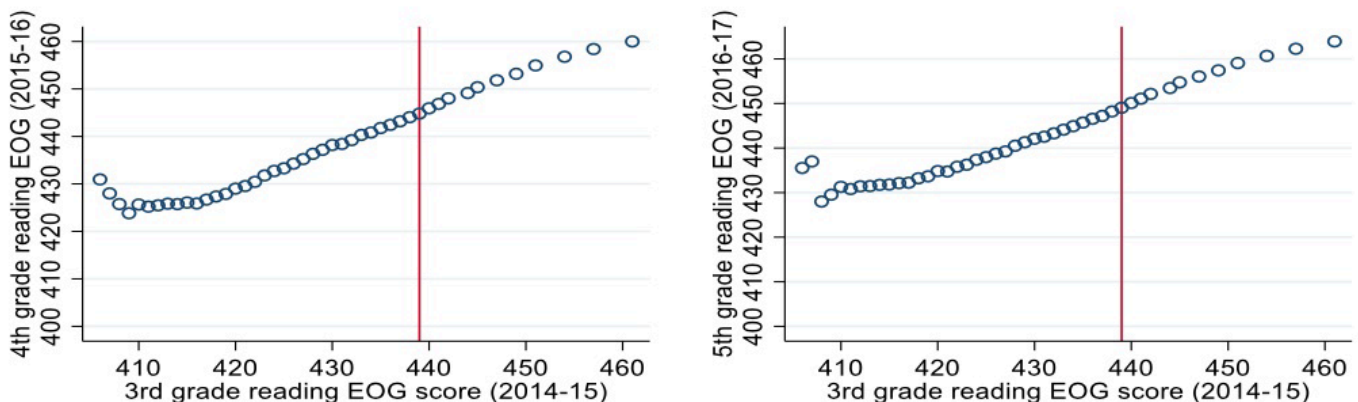


Figure 4. 2013-14 and 2014-15 Cohorts, Two Years Out

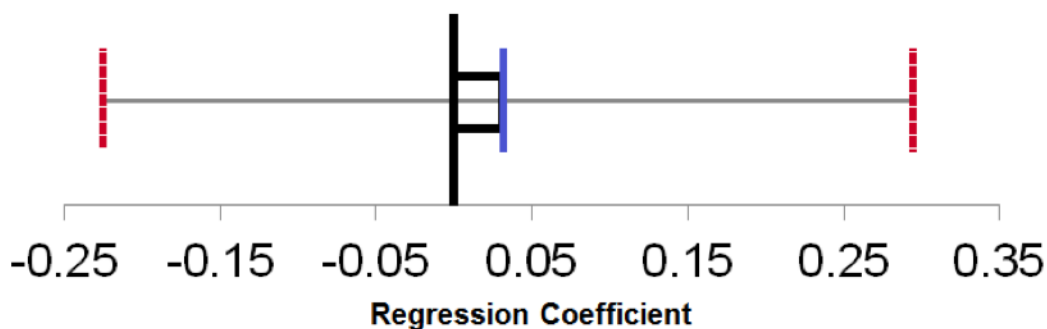


These interpretations of the figures above and of the trends in EOG scores across years in Figure 1 are back-of-the-envelope only, though. To answer the key question—Did impacted students at least do better than they would have without the intervention?—more rigorously, we need to focus on the students whose 3rd grade scores are clustered around the cut-point, for the reasons explained in our outline of our analytic approach.

When we do so, no matter how large a group we look at on either side of the initial cut-point (± 2 points, ± 3 points, or ± 4 points), and no matter whether we look at EOG scores one year later or two years later, the improvements in EOG scores for RtA students are not statistically different from the improvements in EOG scores for students just above the cut-line. Entering the RtA process after completion of the 3rd grade EOG and before enrollment in 4th grade does not appear to have had any impact at the state level on the first two cohorts of RtA students.

As an example, Figure 5 illustrates the outcomes of the statistical analysis for one-year-out test scores for the 2013-14 cohort of 3rd graders. The blue line in Figure 5 represents RtA students' statistically adjusted RtA scores. The thick-lined black rectangle indicates the statistically adjusted estimated difference between RtA students' one-year-later scores and the scores of the students who were just above the cut-line, but this estimate is just that—an estimate—and comes with a margin of error. The red dashed lines indicate the maximum and minimum extent of the margin of error. In order for the RtA students' scores to be meaningfully (that is, statistically significantly) different from the scores of their peers, the space between the maximum and minimum lines cannot include the value 0 (that is, the value that indicates that there is no real difference). In this example, the blue line indicates that the estimated score difference for this group of RtA students is marginally positive (0.03 points), but the margin of error prevents us from being confident that their scores are meaningfully different from the scores of the non-RtA students. Two-year-out results for this cohort (not pictured) are actually marginally lower but also not significantly so; one- and two-year-out results for the 2014-15 cohort are similar (Table C.1, **Appendix C**).

Figure 5. Overall Impact, 1 Year Out (2013-14 Cohort)



Note: These results compare students based on original EOG failure, not eventual placement status

In other words, outcomes for initially-identified RtA students are essentially the same as outcomes for non-RtA students just above the cut-line, whether we look at the first or second cohort, and whether we look at scores one or two years later.

What is the causal effect of being retained in 3rd grade on student reading performance one year later?

Just being among the group of initially identified students only addresses the question of whether RtA as a catch-all policy seems to have any impact. As we indicated earlier, we are not limited to analyses of initially identified students; we also can use the same analytic approach to assess the impact on the subgroup of RtA students who, after a spring and summer of intervention, still are retained at the start of what would have been their 4th grade year and who benefit from at least some exposure to a different classroom structure and curriculum than their peers. When we conduct these analyses, however, there still does not appear to be any significantly different outcome for this group, compared to outcomes from students who just missed identification and received no additional services. Table C.2 in **Appendix C** includes these results.

How do short- and longer-term effects vary by student sub-groups (e.g., gender, race/ethnicity, and economic disadvantage, means of demonstrating proficiency)?

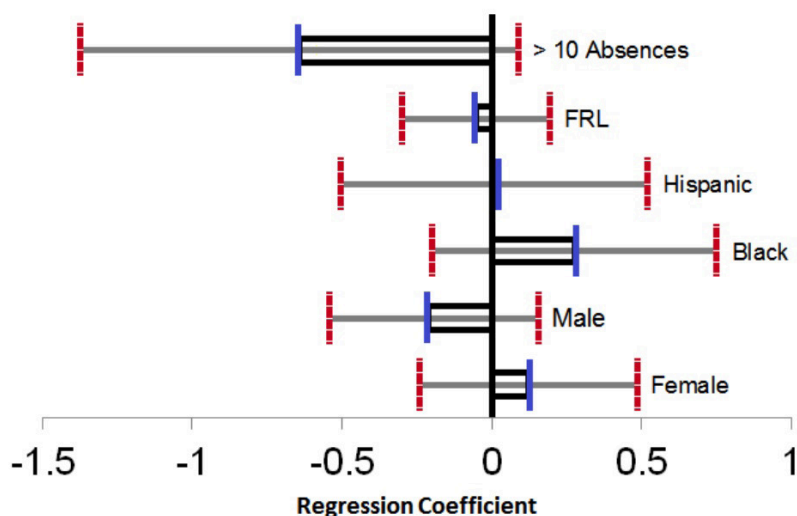
The policy does not appear to be effective for the entire group of either initially identified or eventually retained students, but perhaps it *is* effective for certain sub-groups of students. For example, perhaps lower-income students respond better to the intervention than do others, or perhaps the intervention is more effective for female students. As shown in Table 2, these and other subgroups are subjected to the implications of the policy in different proportions.

Table 2. Initial Identification by Student Subgroup (2013-14 Cohort)

	Passed initial grade 3 EOG (n=ca. 67,000)	Failed initial grade 3 EOG (n=ca. 42,000)
Female	52.0%	46.7%
Male	48.0%	53.3%
Asian	3.9%	1.7%
Black	17.9%	36.7%
Hispanic	12.8%	24.2%
American Indian	1.0%	1.7%
Multi/Other Race	4.1%	3.6%
Pacific Islander	0.2%	0.1%
White	60.1%	32.0%
Free/Reduced-Price Lunch	39.3%	70.2%

Even looking at results for these sub-groups, however (using the same techniques described above)—whether initially identified or eventually retained—there still does not appear to be any effect (Table C.5 and Table C.4, **Appendix C**; example, Figure 6).

Figure 6. Sub-group Impact, 1 Year Out (2014-15 Cohort)



Note: These results compare students based on original EOG failure, not eventual placement status

Does participation in a reading camp make a difference?

One of the interesting features of the program is that a key component—participating in reading camp between the 3rd and 4th grade years—is not mandatory. As a result, each year, a large proportion of students who are eligible for RtA services are not exposed to one of the major interventions. We discuss the implications of this part of the policy later in this report.

To begin to answer questions about the impact of participation in a reading camp, we compared outcomes for students who attended reading camps (both those who attended reading camps and then were promoted, and those who attended reading camps but were retained) to outcomes for students who were eligible for reading camps but did not attend. As was true in all of the other analyses above, however, when assessed at the state level, participation in a reading camp also did not appear to make a difference in subsequent test scores (Table C.5, **Appendix C**).

Policy Implications: Why Are There No Positive Results?

Limitations of the Current Study

By now, many readers likely have picked up on our constantly-repeated caveat, “at the state level.” Most of our analyses to date have been state-level only and therefore would miss any specific local successes. We already have shared our results by student subgroup, and we discuss below results of follow-up analyses we conducted of outcomes based on a student’s placement during the year of retention to help address this deficiency, but neither set of analyses would detect successes at the school or even LEA level. Indeed, we have heard from many practitioners from across the state who believe their localized versions of RtA are having an impact on their students, but because of the sometimes very small size of the group of students impacted in most of the state’s LEAs, we are not able to test these intuitions statistically.

In addition, the strength of the analyses derived from our methodological approach simultaneously introduces a key weakness: Because we have to focus on students whose original EOG reading scores are close to the policy’s eligibility line, we are not able to estimate the impact (using the same methodology) on students whose original scores were much lower—that is, students with much lower reading proficiency at the end of 3rd grade. Our analyses would not pick up whether the subset of students who score several levels below proficiency on their 3rd grade EOGs benefit more from the program than do the students in our analyses.

Two pieces of evidence suggest, however, that, even for these students, the impact has been minor. First, as we noted at the beginning of this report, the overall trend in 3rd and 4th grade EoG scores has been relatively flat every year since the policy was put in place. Second, as shown above in Figure 3 and Figure 4, even without statistical analyses it is relatively easy to see that the average next-year or two-year-out EOG scores for 3rd graders whose scores were well below the cut-line show at best only slightly better improvement than those of their peers, and even then only enough to raise their next-year or two-year-out scores to about the same level of those of their previously slightly-higher-scoring peers.

After thinking through these potential weaknesses of the current study and sharing our quantitative results with several state- and local-level stakeholders, our conclusion is that, rather than being an artifact of the study’s limitations, the general lack of overall progress is more likely attributable to significant gaps between the RtA policy and several aspects of on-the-ground implementation realities.

Potential Policy Challenges

There are four aspects of the original policy that we believe could have contributed to these outcomes.

Lack of Support for Pre-3rd Grade Intervention

RtA began as a 3rd and 4th grade intervention only; our two cohorts—2013-14 and 2014-15 3rd graders—did not have access to any special pre-3rd grade reading interventions other than those offered already to all pre-3rd graders. This challenge has been at least partially addressed in recent years with the addition of screening¹³ and funding earmarked for earlier-grade RtA interventions, but it is beyond the scope of the current study to determine whether or to what extent this additional investment will, in the end, help to improve reading proficiency by 3rd grade. Students in the next two cohorts (3rd graders in 2015-16 and 2016-17) have been at least partial beneficiaries of that extension of the program, and students in last year's cohort (3rd graders in 2017-18) could have had up to three additional years of pre-3rd grade exposure to RtA supports. As we have noted, early returns for later cohorts do not appear to indicate significant improvement, but a definitive conclusion will require additional research.

Broad Definition of Reading Proficiency

Another important challenge is the extensive number of ways in which students either can be excluded from the implications of the policy (via “good cause” exemptions¹⁴) or can be deemed “reading proficient” (via several alternate testing methods; Figure 7, following page). On the one hand, the idea that reading is an isolatable skill with a proficiency level easily determinable by a single test is questionable at best, and the availability of alternate methods of determining a student's true reading proficiency is not just helpful but ultimately necessary. The challenge is less about the availability of alternate assessments and more about whether and to what degree each of those alternate assessments is a valid and reliable tool (i.e., whether some or all of them overestimate students' true reading proficiency).

Assumptions about Availability of Qualified Teach

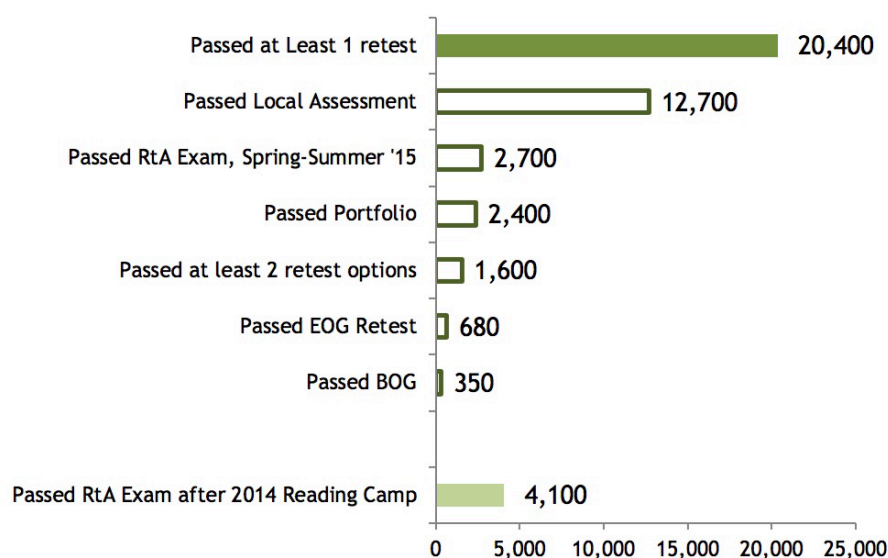
A third potential area of weakness for the overall policy is a result of the state-level lens with which the policy has been applied. This weakness is most readily apparent in the guidelines that require placement of students who *attend reading camps* with “licensed teachers selected based on demonstrated student outcomes in reading proficiency or in improvement of difficulties with reading development,” and placement of students who are *retained* with “a teacher selected based on demonstrated student outcomes in reading proficiency.”¹⁵ In both cases the requirement is pedagogically sound, and most would argue essential; the problem is with the assumption that such teachers are equitably distributed across the state's 115 LEAs, and that enough of those teachers are willing or able to staff the reading camps. Without supporting policies in place to ensure that such teachers are readily available in all elementary schools during the school year and in reading camps over the summer, the requirement is only an aspirational guideline.

¹³ For example, the Kindergarten Entry Assessment (developmental screening) tool now administered to all entering Kindergarten students and the expansion of Reading Camp availability to students in 1st and 2nd grades.

¹⁴ These exemptions include Limited English Proficiency status, status as a Student with a Disability, and previously-remediated and -retained students.

¹⁵ North Carolina General Statutes 115C-83.3 and 115C-83.8

Figure 7. Pathways to Promotion after Initial EOG Failure: 2014-15 Cohort Example



The Nature of RtA Retention

Similarly, the statutory guidelines for the various placement options for retained students (including the transitional 3rd and 4th grade class combination, which is to be “specifically designed to produce learning gains sufficient to meet 4th grade performance standards”¹⁶) also assume that a school housing one or more of these classes has the capacity to truly differentiate the instruction delivered in them. Schools can choose between 3rd grade accelerated reading classes, 3rd and 4th grade transition classes, and 4th grade accelerated reading classes—all with or without additional pull-out reading instruction—but have no guidance to follow with respect to which option(s) might be best for their particular students.

Potential Implementation Issues

All four of the policy challenges outlined above play out in different ways in terms of each LEA’s (and even each school’s) approaches to implementation. There are at least five ways in which these aspects of the original policy may have affected program implementation.

Differences in Local Camp Structure

In 2017, as part of our efforts to learn more about the context in which RtA is implemented across the state, we conducted an informal survey of the representative in each LEA most directly responsible for administering that LEA’s reading camps and asked that representative about her or his LEA’s approach to providing its reading camps. What we learned was that camps vary widely in size (from 400 students to one student, and from 38 teachers to one teacher), length (from the minimum required 72 hours to up to 144 hours), and, most importantly, instructional design (with camps across the state offering various combinations of whole-group instruction, small-group instruction, 1:1 instruction, differentiation, content integration, and other components).

¹⁶ North Carolina General Statute 115C-83.3

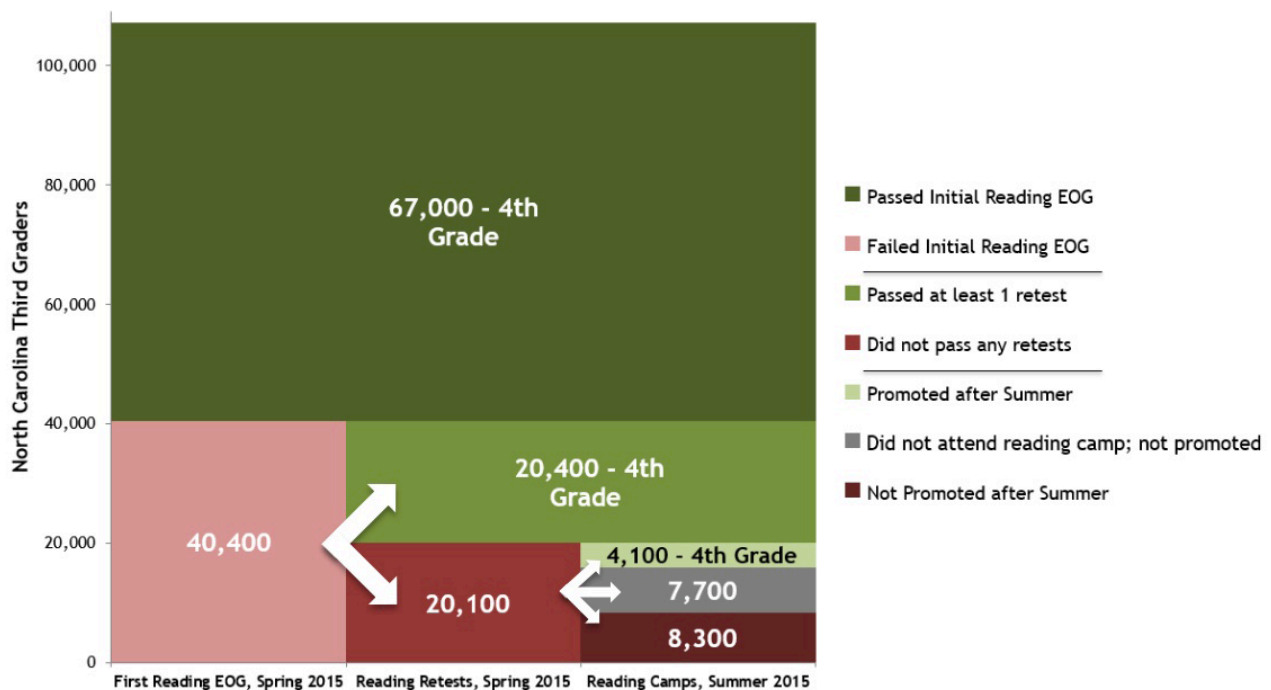
Differences in Local Capacity

As noted above, the legislation assumes equitable availability statewide of high-quality reading teachers (both for school-year classrooms and for reading camps), but in practice such availability is much less consistent. As an example, our reading camp survey revealed the lengths to which some LEAs have to go in order to ensure that their camps are staffed at all—much less staffed with educators who have demonstrated proficiency in reading instruction. While many camps were led by principals, assistant principals, or instructional coaches, 43 of the 110 LEAs who participated in the survey reported that they offered one or more camps led by “other staff,” ranging from classroom teachers to non-instructional support staff. Also, survey results about camp length and conversations with LEA-level implementers indicate that some LEAs have more resources to supplement their provision of camp services than do others. These human resource and fiscal capacity differences likely are true for the school-year component of RtA as well, though measuring the extent of these issues will require additional LEA-level investigation.

Variability in Interventions ahead of the 4th Grade Year

Figure 2 (page 15) demonstrates graphically the proportion of students at each test level below the eligibility cut-line who were not retained for various reasons (including good-cause exemptions). Figure 8 provides more detail about the relative importance of each of the methods for determining proficiency—as well as an indication of at least one way in which students are experiencing different levels of intervention ahead of their 4th grade year.

Figure 8. Retention Attrition, 2014-15 Cohort



After the first EOG test in the spring, a little over 40,000 students initially were eligible for RtA intervention, and over the rest of the spring, about half of those were determined to be proficient via a number of different re-tests—with a large portion of those deemed eligible via a local assessment (Figure 7, page 22). Of the 20,000 students who did not demonstrate proficiency on alternate tests, over 12,000 attended reading camps, but only a little over 4,000 of those demonstrated proficiency at the end of the camp. In addition, almost 8,000 eligible students did not attend reading camps. Not only is proficiency after the spring EOG defined largely by a variety of local assessments, but also a sizeable proportion of those not identified as proficient are not accessing the instructional remediation offered ahead of what would have been their 4th grade year.

Variations in the Retained Student Experience

Interventions in the weeks following initial EOG failure and short-term reading camps collectively can go only so far in accelerating the development of reading proficiency. The majority of the initiative's interventions occur in the 4th grade year, when students who have not demonstrated proficiency can spend up to an entire year in a classroom that is intended to be specifically staffed and designed to address reading challenges. As explained in the **Introduction** and as indicated by the flowchart in **Appendix A**, RtA students potentially can experience very different pedagogical interventions, depending on the options provided by their local schools.

The scope of our analyses did not allow us to investigate this question thoroughly, but we were able to begin the process of investigating differences in the impact of these experiences with the help of data provided by a number of elementary school principals who responded to an informal survey. Through this survey, about one-third of those principals across the state shared with us the various placement options in their schools for RtA students after third grade. In most cases (two-thirds of the schools that responded to the survey), schools provide a single placement option, with most of the remaining schools providing two placement options. Overall (totals add up to more than 100% because of multiple options in some schools):

- 65% of the schools that responded to the survey offer a 3rd grade/4th grade transition class
- 42% offer a 4th grade accelerated reading class
- 25% offer traditional 3rd grade repetition
- 9% offer a 3rd grade accelerated reading class

The outcomes of our informal analysis are reported in Table C.7 in Appendix C. As our analyses of the various pathways suggests, from a statewide perspective, outcomes related to placement in the non-traditional retention pathways (that is, all pathways other than traditional 3rd grade retention) varied. All were positive, relative to traditional 3rd grade retention, and one pathway (placement in a 3rd grade accelerated reading class) even produced a statistically significant outcome; however, our analyses were limited to a handful of students (in fact, only 43 students in the 3rd grade accelerated pathway), and those students were enrolled in a non-random sample of schools. The analyses also are not able to address the critical issue of differences in on-the-ground implementation across schools for any given pathway. We include a recommendation for a more thorough analysis of the impacts of these various placement options in the final section of this report.

Promotion to Grade 5

Finally, the structure of the RtA initiative is such that, rather than being an actual retention policy, it is in practice more like a single-year intervention. The RtA promotion policy applies only to students in 3rd grade; all other promotions—including the potential mid-year promotion after November 1 of the year following 3rd grade—are at the discretion of the principal, per statute.¹⁷ This structure means that, while an individual principal can retain an RtA student for a second year, if she or he still has not demonstrated proficiency, there is no policy requirement to do so.

Our purpose in highlighting this component of the initiative is not to take issue with the idea of principal discretion when it comes to promotion; indeed, the state statute governing all North Carolina promotions acknowledges the importance of principal discretion and personal knowledge of an individual student's true abilities. Our purpose instead is to note that, in practice, almost no RtA students are retained again after what would have been their 4th grade year. In other words, the RtA policy neither ends social promotion, as originally intended, nor does it provide additional supports beyond the single year of implementation, should the students remain non-proficient; instead, the policy introduces a calendar year of potentially intensive interventions for some students identified as non-proficient readers and then returns them to their original trajectories without ongoing support, whether their reading skills have improved or not.

¹⁷ North Carolina General Statute 115C-288

Moving Forward: What the State Can Do Next

The outcomes of the analyses conducted for this report provide little support for continuing the state's current approach to early-grade reading remediation. If there is interest in reforming RtA, the state should consider at least three areas of focus.

Improve Implementation Fidelity

Our reading camp and next-year-placement surveys suggest that there is an overall need for more support to ensure greater implementation fidelity of all components of RtA across all LEAs. In many cases, the lack of fidelity has less to do with local interpretation of the policy and more to do with inequitable distribution of resources—both in terms of human and fiscal capital. For example, what would it take in terms of time, instruction, and staffing to ensure that every camp in the state is able to operate under what developmental reading experts consider to be minimum effective principles? What transportation or other supports should the state provide to ensure that more students are exposed to those camps? And, most importantly, what additional human capital, financial, pedagogical, and material resources are needed in order to strengthen the school-year intervention component of the initiative?

Identify and Scale Up Local Successes

Our surveys also helped us to get a better understanding of the sometimes vast differences in the ways in which the RtA policy is translated at the local level. Because of this implementation diversity, the initiative is in practice essentially 115 different pilot programs operating under a few common parameters. As we noted earlier, our conversations with LEA- and state-level implementers indicate that, at least anecdotally, there are positive things happening in some RtA scenarios. One important question for the state to investigate might be whether it is possible to identify specific implementation combinations that contribute to local positive outcomes.

Our incomplete analyses of outcomes for students based on their participation in a reading camp and their placement the following school year is only the beginning of such an investigation. A more complete study might include deeper consideration of the following components as well:

- Implementation in Kindergarten through 2nd grade
- Determination of proficiency
 - Reconsideration of eligible good-cause exemptions
 - Relative fidelity of each alternate measure of proficiency
- Camp implementation
 - Overall structure
 - Staffing
 - Length
 - Timing

- School-year placement structure and implementation
 - o Instructional models
 - o Staffing
 - o Impact on other subject areas¹⁸

There is always a danger of “over-fishing”—that is, conducting iterative rounds of data-diving in search of a quantifiable “definition” of a successful combination of components—but a commitment to careful, mixed-methods efforts to identify and reach a better understanding of local successes could be time well spent.

Transition from a Social Promotion Mindset to a Literacy Development Mindset

It is outside our professional expertise and beyond the scope of the analyses we were able to conduct for this report to recommend wholesale changes in the philosophical underpinnings of RtA, but we believe the outcomes of our analyses and surveys do suggest that there is not only value in but also the need for a reconsideration of the state’s overall approach to improving early-grade literacy. The rationale for a focus on ending social promotion is clear, but the early evidence is that this focus has not resulted in the original desired outcomes of the policy. To begin the process of making the transition from a focus on social promotion to a focus on literacy development, policy-makers and state and local RtA implementers may benefit from inclusion of a wider representation of North Carolina’s early childhood and literacy experts in planning for the next stages of RtA.¹⁹

¹⁸ We began this process by completing an initial investigation into whether an increased focus on literacy had a negative impact on mathematics instruction (as indicated by declines in math EOG scores one year later) and found no impact.

¹⁹ As an example, the North Carolina Early Childhood Foundation is in the process now of construction recommendations for early reading literacy as part of its Pathways to Grade-Level Reading Initiative Action Framework: <https://buildthefoundation.org/initiative/pathways-to-grade-level-reading/>

References

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1-26.
- Bloom, H.S., Hill, C.J., Black, A.R., and Lipsey M.W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1, 289–328
- Boyd, D., Lankford, H., Loeb, S., Rockoff, J. Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793-818.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82, 2295–2326.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b Robust data-driven inference in the regression-discontinuity design. *Stata Journal*, 14, 909-946.
- Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The long-term impact of teachers: Teacher value-added and student outcomes in adulthood. NBER Working paper 17699.
- Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, 65(1), 1-118.
- Dee, T.S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 43, 528-554.
- Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review*, 14(1), 1-21.
- Gelman, A. and Imbens. G.W. (2014). Why high-order polynomials should not be used in regression discontinuity designs. Working Paper no. 2040, National Bureau of Economic Research.
- Gettinger, M. (1995). Best practices for increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 343–354). Washington, DC: National Association of School Psychologists.
- Goldhaber, D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.

- Goldhaber, D., & Hansen, M. (2009). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Education Research Journal*, 47(1), 218-251.
- Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Evaluating the long-term stability of teacher performance. *Economica*, 80, 589-612.
- Greene, J., & Winters, M. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, 2(4), 319-340.
- Greene, J. & Winters, M. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review*, 28, 135-142.
- Hanushek, E. A., Kain, J. F., O'Brien, D., & Rivkin, S. (2005). The market for teacher quality. NBER Working Paper.
- Harn, B. A., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction: Does additional instructional time make a difference for the most at-risk first graders?. *Journal of Learning Disabilities*, 41(2), 115-125.
- Imbens, G. W., and K. Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79, 933–959.
- Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615–635.
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, 86(1), 226-244.
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420-437.
- Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, 83, 386-431.
- Lee, D., and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655–674.
- Mariano, L. T., & Martorell, P. (2012). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, 35(1), 96-117.
- Marzano, R. J., Gaddy, B. B., & Dean, C. (2000). *What works in classroom instruction*. Aurora, CO: Mid-continent Research for Education and Learning.
- McMaster, K. L., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, 71, 445-463.

- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25, 92-104.
- Range, B. G., Holt, C. R., Pijanowski, J. C., & Young, S. (Spring 2012). The perceptions of primary grade teachers and elementary principals about the effectiveness of grade-level retention. *The Professional Educator*, 36(1). Retrieved from <http://www.theprofessionaleducator.org/>
- Rivkin, S., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94, 247-252.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23, 197-227.
- Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, 27(4), 309-340.
- Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24, 333-357.
- Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher*, 38(2), 120-131.
- Schochet, P.Z. (2008). *Statistical power for regression discontinuity designs in education evaluations*. Technical methods report NCEE 2008-4026, Institute of Education Sciences.
- Schwerdt, G., & West, M. R. (2013). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *IZA Discussion Papers 7314*, Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2250292
- Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research & Practice*, 17(4), 227-241.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S., Chen, R., Pratt, A., Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.
- What Works Clearinghouse (2014). *Procedures and Standards Handbook Version 3.0*.
- Winters, M. & Greene, J. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy*, 7(3), 305-330.

Appendix A. Intervention Flowcharts

Figure A.1. North Carolina Read to Achieve Program End-of-Grade-Three Flow Chart

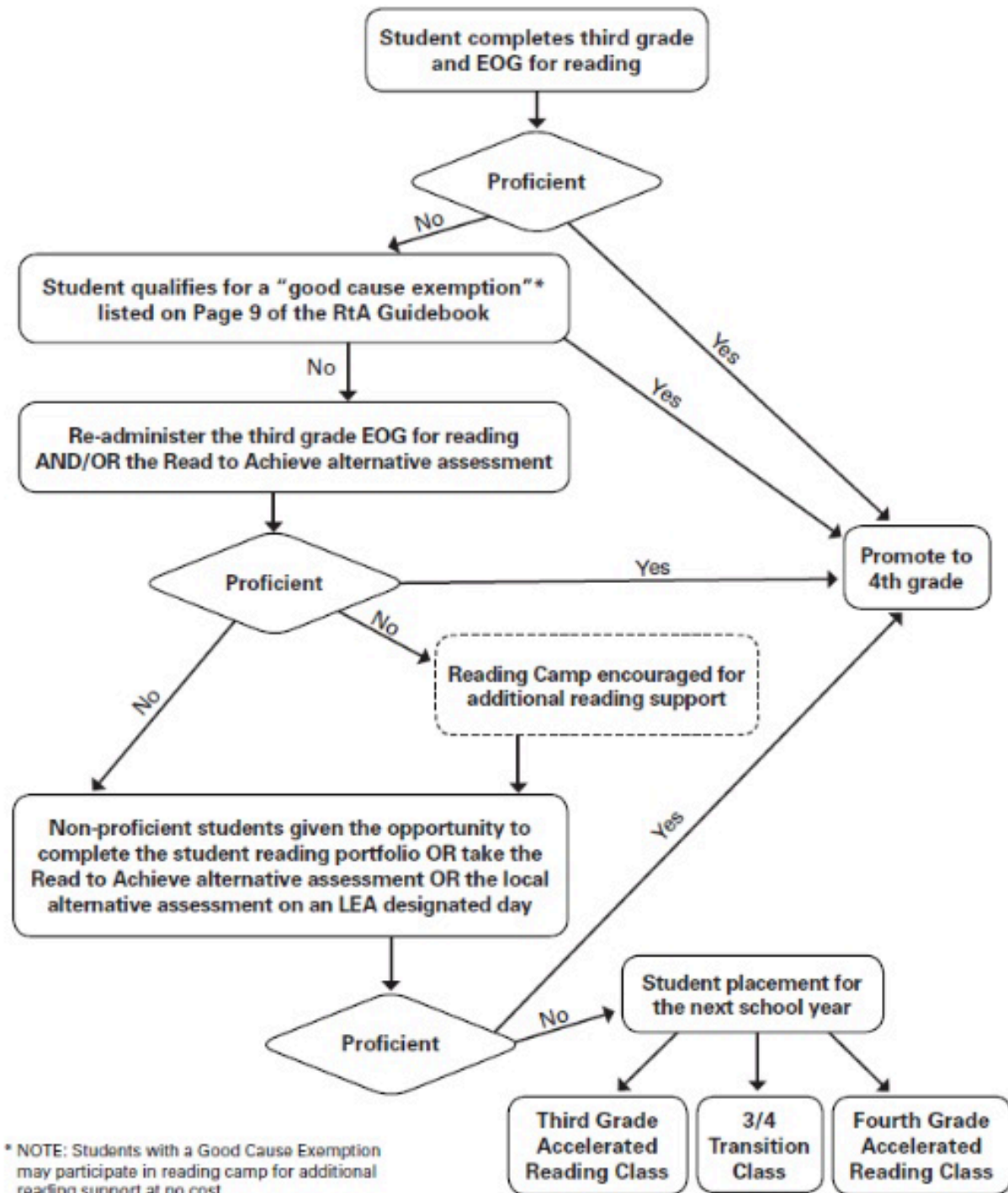
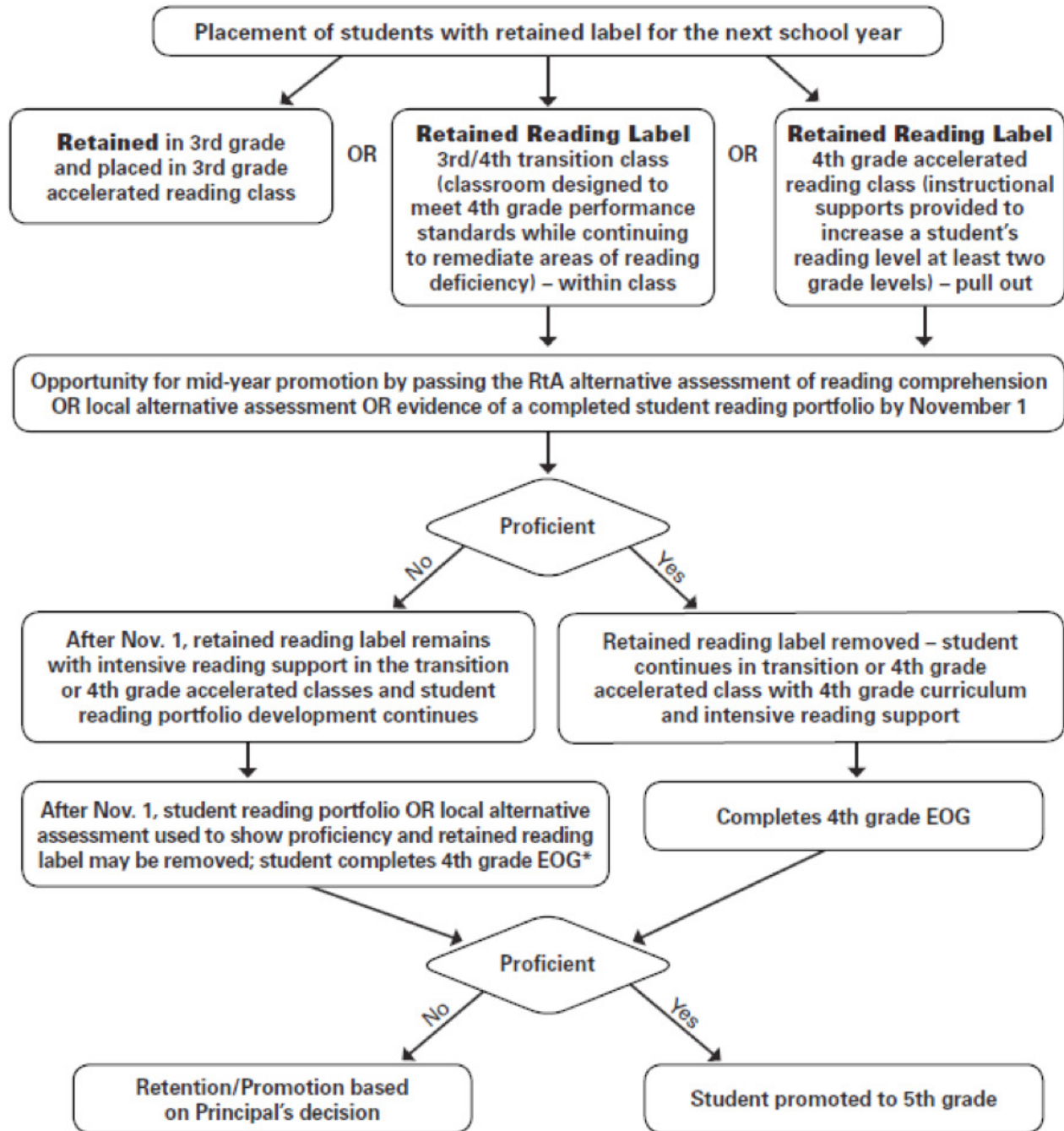


Figure A.2. North Carolina Read to Achieve Program Retention Flow Chart



* After November 1, principal shall grade and classify students demonstrating reading proficiency

Appendix B. Review of Recent Research on the Effectiveness of Read to Achieve Components

Student Motivation and Negative Consequences

The idea that students might be more motivated to learn under the threat of high stakes, such as retention, is theoretically rooted in goal theory (e.g., Ames, 1992; Pintrich, 2000), one aspect of which suggests that students are motivated to avoid negative consequences (such as not being the lowest in the class). There is emerging evidence to suggest that the threat of retention motivates students. One study revealed that elementary teachers believe that the threat of retention increases students' motivation to attend school and parents' motivation to be involved (Range, Holt, Pijanowski, & Young, 2012), while another showed that the threat of retention increases students' motivation to work harder (Roderick & Engel, 2001), but this motivation did not translate into progress and achievement for struggling students. Also, this finding applied to older students (e.g., Bandura, 2001); younger students may not be as sensitive to incentive threats and may lack the agency to give the necessary effort (Roderick et al., 2002).

Grade 3 Retention and Student Achievement

Several districts and states have implemented performance-based retention policies over the past two decades. Early research investigating the impact of elementary and secondary retention on student outcomes suggested that students are harmed by grade repetition (Jimerson, 2001), but identification strategies used in these early studies did not account for potential sources of selection bias. The policies that have received the most attention include district policies in Chicago and New York City and a statewide policy in Florida (Jacob & Lefgren, 2004; Roderick & Nagaoka, 2005; Greene & Winters, 2007; Mariano & Martorell, 2012; Schwerdt & West, 2013). Like North Carolina's RtA, these policies include: a) identification via testing at the end of 3rd grade, with opportunities for re-takes thereafter; b) availability of a summer reading intervention; c) grade repetition if students still do not demonstrate proficiency; and d) some degree of local discretion in student retention. In all three interventions, the policies created a non-linear jump in the probability of grade retention at the achievement-based cutoff, allowing the use of regression discontinuity designs to estimate the causal effect of these policies. In Chicago, New York, and Florida, grade retention improved students' mathematics and reading short-term outcomes; in Chicago, effects faded over time, but they remained stable in the New York and Florida cases (Jacob & Lefgren, 2004; Roderick & Nagaoka 2005; Greene & Winters, 2009; Mariano & Martorell, 2012; Winters & Greene, 2012).

Summer School and Student Achievement

It is difficult to identify separate effects for retention and summer school because positive impact of grade retention might be driven by the increased hours of instruction over the summer, rather than to re-exposure to materials and standards. Jacob and Lefgren (2004) found that the summer school portion of the Chicago retention policy led to an increase of approximately two months of learning for low-performing students, while Mariano and Martorel (2012) found limited evidence that the summer school intervention in New York City's retention program had a significant and lasting impact on student achievement. Meta-analyses of the impact of summer school interventions on student achievement (Cooper et al., 2000; Kim & Quinn, 2013) found that

summer school attendance led to an approximately 0.10 to 0.25 SD increase in student achievement. Both meta-analyses suggest that program characteristics mediate the effectiveness of summer reading programs.

Teacher Quality and Student Achievement

While about two-thirds of the variation in student achievement occurs outside of the school, research consistently demonstrates that teachers play an important role in students' learning trajectories (Chetty et al., 2011; Rivkin et al., 2005; Rockoff, 2004). Research on the impact of teachers on student outcomes documents two important trends. First, researchers consistently find that teachers' observable characteristics (e.g., Master's degree, credential type) have little to no impact on student achievement (Goldhaber & Brewer, 2000; Rivkin et al., 2005; Rockoff, 2004), with the exceptions of teachers' experience, teachers' mathematic and verbal ability, and teacher-student matching (Boyd et al., 2008; Dee, 2007; Ehrenberg & Brewer, 1995; Goldhaber & Hansen, 2009), although the effects are small in magnitude (~.05 SD). Second, unobservable characteristics of teacher quality, often measured using value-added models, have a statistically significant impact on student outcomes. Research on the impact of unobservable teacher quality suggests that a standard deviation increase in a teacher's value-added score leads to a .1 to .2 standard deviation increase in student achievement (Goldhaber & Hansen, 2013; Hanushek et al., 2005; Rivkin et al., 2005).

Additional Reading Instruction and Student Achievement

Actively engaging students in instruction focused on important skills can contribute to achievement (Gettinger, 1995; Marzano et al., 2000). In one study, grade one students who received small-group instruction (four to five students per group) five days per week for 60 minutes outperformed students who received 30 minutes of instruction on measures of word reading and oral reading fluency (Harn et al., 2008). While it is true that researchers who looked at the effectiveness of reading interventions on struggling readers who had responded poorly to previous interventions (McMaster et al., 2005; Vadasy et al., 2002; Vellutino et al., 1996) found that many students did not respond to additional weeks of intervention and continued to experience extreme difficulty learning to read, instructional time for reading also tends to decline in later grades (e.g., 90 minutes per day for grade one; 65 minutes per day for grade five; Rowan & Correnti, 2009), so efforts to increase instructional time for students who are struggling beyond the early primary grades could yield significant effects.

Appendix C. Outputs from Statistical Models

How to Read these Outputs

For most of these tables, the first row is the focus of the analysis (highlighted in grey); the other rows represent the relationship between other factors and our outcome of interest (typically EOG test scores for later years).

There are two sets of numbers for each row: the first number is the estimated impact of the program component of interest (initial eligibility, eventual retention, reading camp attendance, etc.) on test scores either one year or two years later, and the second number (the standard error, in parentheses) helps us determine how large our estimated impact would need to be (no matter whether it is positive or negative) in order to be considered “meaningful” (see definition, below) and not just the result of random fluctuation. The larger the standard error, the larger an estimated impact needs to be in order to be classified as meaningful.

Meaningful, or *statistically significant*, results are marked with either a ** (95% confidence that the result is meaningful) or a *** (99% confidence that the result is meaningful). It is not at all uncommon in these tables to see that a student’s 3rd grade test score is the most significant predictor of her or his score the next year.

Keep in mind that, even if a result is *statistically* meaningful, it may not be *practically* meaningful if the estimated impact is small.

Results of the Sharp Regression Discontinuity Analyses

Table C.1. *Student Reading Performance One Year and Two Years Later (All Initially Impacted Students, Cohorts 1 & 2)*

	2013-14		2014-15	
	<i>1 yr. later</i>	<i>2 yrs. later</i>	<i>1 yr. later</i>	<i>2 yrs. later</i>
Below cutoff	0.0315 (0.1381)	-0.0156 (0.1449)	-0.0461 (0.1356)	-0.0736 (0.1445)
3 rd grade EOG	1.0769*** (0.0367)	1.0626*** (0.0386)	1.0545*** (0.0361)	1.0303*** (0.0384)
Below*EOG	-0.1995*** (0.0578)	-0.1944** (0.0606)	-0.2914*** (0.0569)	-0.2122*** (0.0606)
Constant	444.9994*** (0.0725)	448.9798*** (0.0761)	444.8125*** (0.0710)	449.0512*** (0.0755)
<i>n</i>	32,353	31,379	33,260	32,312

Results of the Fuzzy Regression Discontinuity Analyses

Table C.2. *Reading Performance One Year Later for Retained Students (Cohorts 1 & 2)*

	2013-14	2014-15
Retained	4.4012 (2.9343)	3.2558 (2.1810)
EOG score	1.0776*** (0.0381)	1.0560*** (0.0375)
Retained*EOG	-0.7959 (0.6083)	-0.9217 (0.4633)
Constant	444.9987*** (0.0752)	444.8110*** (0.0738)
<i>n</i>	32,321	33,248

Results for Subgroups

Table C.3. Student Reading Performance One Year and Two Years Later (All Initially Impacted Students, by Subgroup)

	<i>Female</i>	<i>Male</i>	<i>Black</i>	<i>Hispanic</i>	<i>FRL</i>	<i>Absent</i>
2013-2014, 1 yr. later						
Below cutoff	-0.2560 (0.1915)	0.3181 (0.1992)	0.0193 (0.2573)	-0.1619 (0.2906)	0.1146 (0.1679)	-0.0852 (0.4264)
3 rd grade EOG	1.0599*** (0.0505)	1.0939*** (0.0535)	1.1536*** (0.0741)	0.9508*** (0.0824)	1.0401*** (0.0474)	1.1674*** (0.1169)
Below*EOG	-0.2458** (0.0804)	-0.1627 (0.0830)	-0.4233*** (0.1091)	-0.1191 (0.1230)	-0.1684 (0.0709)	-0.3242 (0.1796)
Constant	445.2590*** (0.0997)	444.7340*** (0.1053)	443.9221*** (0.1419)	444.5153*** (0.1577)	444.3931*** (0.0910)	444.6394*** (0.2257)
<i>n</i>	16,224	16,097	8,906	6,063	20,435	3,434
2013-2014, 2 yrs. later						
Below cutoff	-0.1925 (0.1999)	0.1596 (0.2098)	0.0386 (0.2655)	-0.2004 (0.3022)	-0.0359 (0.1770)	-0.3955 (0.4587)
3 rd grade EOG	1.0436*** (0.0527)	1.0799*** (0.0564)	1.0122*** (0.0766)	1.0722*** (0.0858)	1.0285*** (0.0501)	1.1329*** (0.1253)
Below*EOG	-0.2174** (0.0839)	-0.1740 (0.0876)	-0.1956 (0.1128)	-0.3325** (0.1280)	-0.2151** (0.0748)	-0.4062 (0.1932)
Constant	449.1497*** (0.1042)	448.8067*** (0.1111)	447.6277*** (0.1465)	448.9162*** (0.1640)	448.3520*** (0.0960)	448.4641*** (0.2416)
<i>n</i>	15,771	15,601	8,632	5,949	19,933	3,310
2014-2015, 1 yr. later						
Below cutoff	0.1250 (0.1896)	-0.2122 (0.1940)	0.2768 (0.2518)	0.0152 (0.2780)	-0.0519 (0.1738)	-0.6427 (0.4325)
3 rd grade EOG	1.0939*** (0.0500)	1.0155*** (0.0521)	1.0347*** (0.0721)	0.9371*** (0.0783)	0.9794*** (0.0491)	0.9725*** (0.1165)
Below*EOG	-0.3000*** (0.0796)	-0.2836*** (0.0813)	-0.2760 (0.1071)	-0.1597 (0.1173)	-0.2411** (0.0738)	-0.3553 (0.1810)
Constant	444.7907*** (0.0993)	444.8323*** (0.1015)	443.7920*** (0.1381)	444.7166*** (0.1488)	444.2120*** (0.0937)	444.6201*** (0.2235)
<i>n</i>	16,558	16,691	9,217	6,413	18,940	3,496
2014-2015, 2 yrs. later						
Below cutoff	-0.1882 (0.2005)	0.0217 (0.2079)	0.2484 (0.2656)	-0.0409 (0.2984)	-0.1358 (0.1864)	-1.3518** (0.4686)
3 rd grade EOG	1.0071*** (0.0529)	1.0490*** (0.0558)	1.0253*** (0.0761)	0.9632*** (0.0839)	0.9511*** (0.0526)	0.9890*** (0.1262)
Below*EOG	-0.2183** (0.0842)	-0.2137 (0.0871)	-0.2096 (0.1130)	-0.1487 (0.1260)	-0.1722 (0.0791)	-0.5615** (0.1958)
Constant	449.2577*** (0.1050)	448.8495*** (0.1085)	447.8374*** (0.1455)	448.9313*** (0.1591)	448.3816*** (0.1003)	448.7533*** (0.2427)
<i>n</i>	16,115	16,187	8,951	6,256	18,438	3,367

Table C.4. Reading Performance One Year Later for Retained Students (by Subgroup)

	<i>Female</i>	<i>Male</i>	<i>Black</i>	<i>Hispanic</i>	<i>FRL</i>	<i>Absent</i>
2013-14						
Retained	-0.7930 (3.8147)	10.2062 (4.5643)	5.1530 (4.2495)	-1.0472 (4.6647)	5.0211 (3.2621)	5.0225 (9.6708)
EOG score	1.0603*** (0.0515)	1.0938*** (0.0568)	1.1536*** (0.0815)	0.9508*** (0.0830)	1.0401*** (0.0492)	1.1674*** (0.1243)
Retained*EOG	-1.7143 (0.7851)	0.1500 (0.9550)	-1.8076 (0.9030)	-0.9027 (0.9553)	-0.3262 (0.6503)	-1.2298 (1.8656)
Constant	445.2587*** (0.1019)	444.7339*** (0.1118)	443.9221*** (0.1561)	444.5153*** (0.1588)	444.3930*** (0.0944)	444.6394*** (0.2401)
<i>n</i>	16,223	16,096	8,906	6,063	20,434	3,434
2014-15						
Retained	6.7102 (3.3078)	0.1733 (2.8933)	7.3622 (3.7997)	1.9016 (3.5705)	2.1620 (2.4693)	-5.2544 (5.1540)
EOG score	1.0939*** (0.0528)	1.0155*** (0.0534)	1.0347*** (0.0771)	0.9376*** (0.0798)	0.9794*** (0.0505)	0.9725*** (0.1180)
Retained*EOG	-0.3543 (0.6923)	-1.3950 (0.6250)	-0.0327 (0.7750)	-0.3379 (0.7408)	-0.7025 (0.5056)	-2.5412 (1.2328)
Constant	444.7907*** (0.1052)	444.8323*** (0.1040)	443.7920*** (0.1476)	444.7172*** (0.1517)	444.2120*** (0.0966)	444.6201*** (0.2264)
<i>n</i>	16557	16691	9217	6412	18940	3496

Results for Reading Camp Participation

Table C.5. Student Reading Performance One Year Later for Participants in Reading Camps

	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>
Attended camp	-0.9029 (2.5484)	-1.2087 (3.0514)	-2.5353 (2.7868)
Age		-1.2575*** (0.1496)	-1.2082*** (0.1508)
Female		0.8208*** (0.1443)	0.7672*** (0.1444)
Black		-1.5841*** (0.3195)	-1.4150*** (0.2681)
Hispanic		-0.5063 (0.2537)	-0.4558 (0.2581)
FRL		-1.4328*** (0.1897)	-1.5542*** (0.2075)
Days absent		-0.0625** (0.0239)	-0.0752*** (0.0222)
Constant	435.3625*** (1.8651)	448.6816*** (2.7754)	452.6342*** (3.3440)
<i>n</i>	10,264	10,206	10,206

Notes: Instrument for camp attendance is travel time in minutes between residence and nearest district summer camp. Dependent variable is reading EOG one year after initial 3rd grade EOG. Model 3 includes district fixed effects (coefficients not shown).

Results Based on Classroom Placement Type for Retained Students

Principals in about 400 (of about 1,300) elementary schools responded to our request to describe the placement options offered at their schools for students retained as a result of RtA. As a result, the analyses in this section are for a non-random subset of all impacted students; we do not draw firm conclusions from these results and include them primarily to inform initial discussions about the relative impact of various placement options.

Table C.6. *Identified Placement Options and Number of Students Enrolled in Each, Survey Participant Schools Only*

<i>Classroom Type</i>	<i>Students</i>	
	<i>(n)</i>	<i>%</i>
<i>Single Placement Option</i>		
3 rd /4 th transition class	1,438	41.8
4 th grade accelerated reading class	822	23.9
Traditional 3 rd grade retention	51	1.5
3 rd grade accelerated reading class	43	1.3
<i>Multiple Placement Options</i>		
Traditional 3 rd & 3 rd /4 th transition	262	7.6
Traditional 3 rd & 4 th grade accelerated	236	6.9
3 rd /4 th transition class & 4 th grade accelerated	197	5.7
Other multiple placement combinations	394	11.4
Total	3,443	

Table C.7. Student Reading Performance One Year Later, by Classroom Placement Type(s) in Responding Schools

Classroom Type	
<i>Single Placement Option</i>	
3 rd /4 th transition class	1.7003 (1.2609)
4 th grade accelerated reading class	2.4059 (1.2993)
3 rd grade accelerated reading class	5.1582** (1.9193)
<i>Multiple Placement Options</i>	
Traditional 3 rd & 3 rd /4 th transition	1.3514 (1.3526)
Traditional 3 rd & 4 th grade accelerated	1.3043 (1.3674)
3 rd /4 th transition class & 4 th grade accelerated	3.2660 (1.3524)
Other multiple placement types	2.2739 (1.3305)
FRL	-1.4695*** (0.3208)
Days absent	-0.0401 (0.0236)
Female	0.4780 (0.2496)
Age	-1.1376*** (0.2619)
Asian	1.4813 (1.2349)
Black	-1.1792*** (0.3426)
Hispanic	-0.1931 (0.3423)
Constant	442.3566*** (2.7667)
<i>n</i>	3,232

Notes: Standard errors in parentheses; model includes district fixed effects (coefficients not shown); Reference category for placement indicators is Traditional 3rd grade class; Dependent variable is reading EOG one year after initial 3rd grade EOG

Contact Information:

Please direct all inquiries to Trip Stallings

dtstalli@ncsu.edu