CAN WE MOVE FORWARD BY HOLDING BACK?: AN ANALYSIS OF THE EFFECTS OF TEST-BASED GRADE RETENTION IN THIRD GRADE ON STUDENT ACHIEVEMENT IN INDIANA

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF EDUCATION

BY

VINCENT A. EDWARDS

DISSERTATION ADVISOR: SERENA SALLOUM

BALL STATE UNIVERSITY

MUNCIE, IN

JULY 2021

ABSTRACT

DISSERTATION: Can We Move Forward By Holding Back?: An Analysis of the Effects of Test-Based Grade Retention in Third Grade on Student Achievement in Indiana

STUDENT: Vincent A. Edwards

DEGREE: Doctor of Education

COLLEGE: Ball State University Teachers College

DATE: July 2021

PAGES: 121

As of 2021, eighteen states plus the District of Columbia have implemented legislation regarding mandatory test-based grade retention focusing on minimum reading proficiency levels (National Conference of State Legislatures, 2019) with many of those policies targeting the third-grade year. While grade retention has been largely viewed negatively in the research conducted prior to NCLB, some studies of the academic effects of other city- or state-wide test-based retention policies in recent years have found that there can be positive outcomes. Indiana is one of the states with a retention policy tied to IREAD-3, a reading test in Grade 3, but up to now, the effectiveness of this policy has not been studied thoroughly. The decision to retain a student in grade is complex and critical with a variety of factors to understand as well as meaningful and far-reaching consequences to consider. This study focuses on the academic effects of test-based retention on students in the 2012 Grade 3 cohort in Indiana and follows them through Grade 8 using ISTEP+ results. A comprehensive, statewide dataset is analyzed using descriptive and inferential statistical approaches, culminating with longitudinal multilevel modeling to control for many student-level factors as well as their school districts.

Consistent with some of the similar studies conducted in other cities and states, this study finds that retention has a positive effect on academic performance in a same-grade comparison in both English / Language Arts and Mathematics. These findings support the idea that grade retention could be considered a viable intervention policy; however, it is important to consider the ancillary supports and efforts that often accompany retention policies and view the positive gains as an aggregate of these efforts. Considered as a part of the existing body of research on grade retention, there is still much to be understood about most appropriate uses of the practice of grade retention.

Keywords: grade retention, IREAD-3, ISTEP+, high-stakes testing, student academic performance, school accountability, education policy

DEDICATION

To all educators who are relentlessly dedicated to making a difference for their students. I hope this study contributes to understanding how to do what is best for kids.

ACKNOWLEDGMENTS

To paraphrase what many have said before, it is the journey and not the destination. Time, space, and memory make it impossible for me to come up with a list of all of the people who have helped me professionally or personally along the journey to the completion of this dissertation and doctoral degree. Nevertheless, it is with a profound sense of gratitude that I extend a sincere "Thank you!" to some key players.

I am truly grateful for the excellent professors that have shared their time, expertise, and care in my educational experiences of the past few years. First among these, I would like to thank Dr. Serena Salloum, my dissertation chair, for her guidance and encouragement. Also, to all of my dissertation committee, Dr. Marilynn Quick, Dr. Holmes Finch, and Dr. Nick Elam, I would like to thank you for your input and encouragement, and especially to Dr. Finch for his guidance for the statistical analysis. I feel very fortunate to have had a truly stellar committee.

To Dr. Kianre Eouanzui and Nick Mundell for their assistance with the statistical analysis and using SPSS. Their support and advice were extremely valuable.

To my fellow students in the two different cohorts that I primarily went through the program alongside. The entire experience was better due to the company kept along the way.

To all of my extended family for their love and unconditional support. I am truly blessed to have them in my life. To my children, Graham and Marcus, for their support and patience while enduring times when I was less available or more tired.

Finally, and most importantly, I thank my wife, Stacey, for her willingness to be an aider and abettor through this entire process. I would not have made it through to the end of this journey without her love and support.

Soli Deo gloria!

CHAPTER 1

INTRODUCTION

No one wants to fail. However, what we do as human beings when confronted with failure is both revealing and highly consequential. In some cases, such as when attempting to bake a loaf of bread, the stakes are low and results are easily rectified, often by simply doing it over. In other instances, the stakes are high and the consequences irreversible. The education of our youth has, rightly so, become increasingly viewed as a high-stakes endeavor. In recent decades in the United States, with the proliferation of high-stakes student achievement testing as a part of an educational environment emphasizing standards and accountability, educators are continuously confronted by a greater need than ever for research-based decision-making. As state and federal policymakers shape the agendas of our nation's public schools, it must be remembered that the success of individual students in becoming happy and productive citizens is at the heart of public education. It is natural to believe that, when those individual students fail to succeed, it is incumbent on schools to intervene. One consistently used intervention strategy for students who are not able to successfully achieve mastery of the prescribed grade-level curriculum is grade retention (Bowman, 2005; Rose & Schimke, 2012; Smalls, 1997). Grade retention refers to the practice of having a student retained to repeat a grade level instead of being promoted on to the next grade and is at times referred to as failing/flunking a grade, being held back, repeating a grade, or non-promotion.

Reading, in particular, is considered an especially important indicator of future academic success (Butler, Marsh, Sheppard, & Sheppard, 1985; Rose & Schimke, 2012), and an oft-targeted key location in the reading curriculum is prior to fourth grade, which is a grade level many consider the transition from "learning to read" to "reading to learn" (Schwerdt, West, &

Winters, 2017; Winters, 2019). Consequently, as of 2021, eighteen states plus the District of Columbia have implemented legislation regarding mandatory test-based grade retention focusing on minimum reading proficiency levels (National Conference of State Legislatures, 2019) with many of those policies targeting the third-grade year. An additional 11 states have passed legislation that allows for but does not require grade retention based on a reading proficiency test (a list of states with reading retention policies can be found in Appendix B). Since the passing of Public Law 109 in 2010, Indiana has used the IREAD-3 assessment as a measure of minimum reading proficiency for all third graders. Based on this policy, students who do not pass this test, and do not qualify for a "good cause exemption" due to a disability, are presumably retained in third grade. In some cases this means full grade retention, while in other cases school leaders will have the student go on to fourth grade and make them "double up" on reading instruction by retaking third grade reading as a fourth grader to satisfy the particulars of the policy.

Background

Many of the efforts in educational research are focused on understanding all that affects student achievement. In taking responsibility for the achievement of their students, schools have attempted to intervene when students are not performing at a level considered to be sufficient through a wide variety of avenues and strategies. These involve standards-based curriculum, differentiated instructional strategies, multi-tiered systems of support, and various types of assessment. Included in these attempts to individualize pacing and instruction is the practice of grade retention. With the more recent trends to emphasize data-driven instruction and accountability through high-stakes testing, the practice of test-based grade retention has experienced a resurgence. The question facing school administrators and other policymakers is whether or not test-based retention is effective.

As researchers have sought the answer to this question, some have focused on the socioemotional effects on students (Jimerson, 2001; NASP, 2011) while others have targeted solely the outcomes of student achievement levels (Hattie, 2012; Holmes & Matthews, 1984; Schwerdt, West, & Winters, 2017) and still others have looked at both (Marsh, Parker, Guo, Dicke, Pekrun, Murayama, & Lichtenfeld, 2017) or an altogether different mixture of outcomes to families, communities, etc. (LiCalsi, Ozek, & Figlio, 2019). Within the study of how grade-retention affects student achievement, some studies have solely examined short-term results within the year or two immediately following the decision to retain (Jacob & Lefgren, 2004) as that proximity to the decision is considered to provide the best measure of the effectiveness of the policy. This line of thinking emphasizes the assumption that the farther away from any intervention, the more confounding variables come into play. Other studies have taken a longer view of the continuing or lasting effects of grade retention and analyze multiple years past such a decision or even focus on graduation or high school dropout rates (Alexander, Entwisle, & Dauber, 2003; Hughes, West, Kim, & Bauer, 2018; Jacob & Lefgren, 2009; Stearns, Moller, Blau, & Potochnick, 2007). This is arguably a more pragmatic and realistic measure of the effectiveness of grade retention in giving the student the "gift of time." That is one of the most frequently made arguments made by proponents of grade retention who see it as giving additional time for development and mastery to students who are struggling. The good intentions of advocates for test-based grade retention are that this extra year will allow the student to catch up and flourish moving forward. As school leaders and policymakers, it is vital to consider the effectiveness of any policy or practice, and if providing an additional year of time to a student does not deliver lasting results beyond the gains stemming from repetition of the

curriculum in a given year, the perceived value of grade retention as an intervention for failing students drastically changes.

Another issue in determining the effectiveness of a test-based retention policy is the simple fact that only one of the two potential avenues is observed (Sekhon, 2007). Either the student is retained or they are not—it is not possible to observe both. Also, since student achievement is so complex an issue and affected by so many confounding variables, it is very difficult to identify homogeneous groups of students for whom it is possible to observe and compare the effect of grade retention as an intervention versus being promoted as normal.

Statement of the Problem

The practice of grade retention affects a significant portion of the student population. According to the U.S. Department of Education, 1.9% of all students in the United States were retained in 2016, and interestingly, although accountability measures and retention policies have increased in across the nation in the same period, this percentage has been slowly declining since 3.1% of all students were retained in 2000 (Status and Trends in the Education of Racial and Ethnic Groups from National Center for Educational Statistics). The policy of grade retention affects not just individual students, but it also has a substantial impact on school systems and their budgets. Using data from the United States Department of Education's National Center for Educational Statistics website (n.d.) and rounding the national average cost to $12,000 per year for a public-school student in the United States, if 2% of the approximately 50 million students in the nation are retained each year, the practice of grade retention costs our national, state, and local governments a combined total of $12 billion each year (Valbuena, Mediavilla, Choi, & Gil, 2020). Since grade retention also has a negative effect on graduation rates (Holmes, 1989; Jimerson, 2001), the subsequent loss to society could be even larger than simply delaying

graduation. With such substantial costs related to the use of this approach to intervention, it is crucial for educational leaders to understand if grade retention is a truly effective policy instrument for the failing students that are being targeted.

Purpose of the Study

The purpose of this study is to analyze the effectiveness of test-based grade retention as a policy and, specifically, to analyze the effectiveness of the IREAD-3 policy in the state of Indiana. The focus of this analysis is on students' academic achievement rather than psychological or social effects. Student testing data from the Indiana Department of Education will be used to examine the effects of grade retention in third grade in accordance with the IREAD-3 policy on the achievement levels of those students on the annually administered Indiana Statewide Testing for Educational Progress Plus (ISTEP+) in the years after the decision to retain the student in grade. Test data from Grades 3-8 will be used in analysis. The independent variables are the IREAD-3 score and the decision to retain or not. The dependent variables are ISTEP+ scores in Grades 4-8. The following variables will be controlled for as they have been shown to be associated with student achievement levels as measured by standardized tests: free/reduced lunch status (a measure of socioeconomic status), IEP (Individualized Education Plan) / disability status, ELL (English Language Learner) status, ethnicity, and gender.

Research Questions

While this study contributes to an understanding of the effectiveness of test-based retention policies in general, the specific research questions of this study relate to effects on student academic achievement of the IREAD-3 policy in the state of Indiana. This study has one

primary research question and two ancillary questions that will directly inform the primary question.

> Primary Research Question: What is the relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?
>
> Subquestion 1: What is the relationship between grade retention and student achievement at each grade level 4-8 after the decision to retain?
>
> Subquestion 2: What is the relationship between grade retention and student achievement considered longitudinally across grade levels 4-8 after the decision to retain?

The choice to use six years after the decision to retain arises from the current grade level and the data available. Six years after the decision to retain would place the student in 8th grade which is on the cusp of entering high school and is also the last year of continuous achievement data from the ISTEP+ assessment.

Theoretical/Conceptual Framework

In this chapter, I briefly address both a theoretical framework and a conceptual framework for this study, which are, in order, Social Ecological Systems theories and the Rubin Causal Model (more details on these frameworks is provided in Chapter 2). Social Ecological Systems theory has been put forward in many different incarnations, but a key figure in its development is Urie Bronfenbrenner (1977, 1979, 1994, 1995, 1999; Bronfenbrenner & Ceci, 1996; Bronfenbrenner & Morris, 1998, 2006). While his theory evolved from the late 1970s until his death in 2005, there was always an emphasis on the ecological aspect of interactions between individuals and the contexts in which they live (Tudge, Mokrova, Hatfield, & Karnik, 2009). McLeroy, Bibeau, Steckler, and Glanz (1988) propose a variation on Bronfenbrenner's

ecological model in which they view human behaviors as influenced by nested levels of factors. These factors will be discussed further in the next chapter and will serve as a lens through which the discussion of findings and implications will be presented.

A desire to find a causal link is at the heart of much of the research performed in the field of education, so it is crucial for researchers to distinguish between correlation and causation. While double-blind randomized experiments are considered the "gold standard" of research (Frank, Maroulis, Duon, & Kelcey, 2013), this kind of research is often impossible or even unethical when it comes to students. For this study, the dichotomy of retained or not retained is a factor that it would be inappropriate to manipulate for research purposes. Rubin's conceptual framework of causal inference, also referred to as the Rubin Causal Model, looks at the differences between potential outcomes, which in this case is two groups with different treatment factors – retained or not – and their average treatment effect to draw conclusions on causality. This framework guides the analysis presented in this study as every attempt is made to isolate the preceding decision of retention in the final statistical model used for analysis.

Significance of the Study

The study contributes to several larger areas of debate for school administrators and policymakers in addition to its direct significance to Indiana and its legislature, educators, and students. This study fits into the body of knowledge on the use of high-stakes testing in education as well as the interventions a school might implement when a student is failing. This study also adds to the research on grade retention. In the significant amount of research over many decades that exists on the practice of grade retention, much of that research centers on the personal effects rather than or in addition to academic on students (e.g. Holmes & Matthews, 1984; Jimerson, 2001; NASP, 2011). Additionally, many studies do not differentiate between

teacher-based grade retention and test-based grade retention, and some researchers argue that this must be done to accurately evaluate the practice (Allensworth & Nagaoka, 2010; Greene & Winters, 2007; Penfield, 2010). While several researchers have examined specific test-based retention policies in recent years (Jacob & Lefgren, 2009; Lorence, 2014; Roderick & Nagaoka, 2005; Schwerdt, West, & Winters, 2017), including similar statewide reading test-based retention policies in other states, there has not been a study on the effectiveness of the IREAD-3 policy in Indiana.

Assumptions and Delimitations

Since I obtained the student testing data from the Indiana Department of Education, I assumed that the data were accurate in relation to specific values and individual students. It was also assumed that the ISTEP+ passing rates are an accurate measurement of student achievement levels. While there are numerous concerns over the use of single-point high-stakes tests (Burger & Krueger, 2003; Campbell, 1979; Geiser & Santelices, 2007; Heubert & Hauser, 1999; Kohn, 2000; Smith & Fey, 2000; Worthen & Spandel, 1991), the use of other state assessments to measure the progress made by students who were retained based on a state assessment seems to fit logically. Additionally, it is assumed that schools in Indiana follow the guidance provided by the IDOE on the IREAD-3 retention policy.

The delimitations of this study include exclusion of some student scores based on incomplete data and/or propensity score analysis. This exclusion is done to limit the comparison of students who were retained or promoted to a sufficiently homogeneous group. This study is additionally limited by the lack of additional information about the processes and deliberations that went into the choice of whether to retain an individual student.

Definition of Terms

The terms included below can be found throughout this study and review of literature.

Cut Score: Also at times referred to as a cutoff-score or minimum passing score, a cut score serves as the hard line between passing and not passing an assessment. Any score equal to or greater than the cut score is considered passing and there is typically little consideration of growth when a cut score is involved (Linn, Betebenner, & Baker, 2002).

High-Stakes Testing: When significant decisions are made for students, teachers, schools, or districts based on student performance on an assessment, typically for the purpose of accountability, the assessment is considered to be high-stakes. These decisions can include student promotion/retention, educator pay levels and job security, budget allocations, public perception of quality, and local control by school boards. This arises from a desire to incentivize higher standardized test scores (Hout, Elliott, & Frueh, 2012).

Indiana Department of Education (IDOE): The administrative entity, in tandem with the Indiana State Board of Education, that provides state-level oversight of public and private schools in Indiana (IDOE, n.d.). The department administers all state assessments and maintains the data produced on students, educators, schools, and districts.

IREAD-3: The Indiana Reading Evaluation and Determination assessment is a test of minimum proficiency in reading given each spring to all third graders in the state of Indiana created as a part of the House Enacted Act 1367 (Public Law 109 in 2010). As a part of the law, student are required to be retained in third grade if they do not receive a passing score on the assessment. There is a provision for one retake in the summer and a good cause exemption for students with a disability. The criterion-referenced assessment produces scores based on Item Response Theory (IRT) with range of 200-650 and a pass cut score of 446 (IDOE, n.d.). In 2019

the overall statewide pass rate on IREAD-3 was 87.3%, and the 2018 pass rate was 87.1% (IDOE, n.d.). Additional information about IREAD-3 pass rates and Good Cause Promotion Exemption rates can be found in Appendix C.

ISTEP+: The Indiana Statewide Testing for Educational Progress Plus assessment has been through many iterations since its initial release in 1988. It is a criterion-referenced minimum proficiency assessment that was replaced by ILEARN in 2019. From 2009-2018, students in grades 3-8 took the tests in the spring of each academic year. The Graduation Qualifying Exam for Indiana High School Diplomas has changed assessments multiple times, and is currently referred to as ISTEP+ 10 . Through Spring 2014, there was an alternative assessment for some students with IEPs called IMAST which was scaled the same way as ISTEP+ but was modified to be more easily passed (IDOE, n.d.).

Social Promotion: The practice of promoting a student to the next grade level along with age peers regardless of demonstrated academic mastery or proficiency is known as social promotion. This practice became the norm in our nation's schools as priorities changed from merit to efficiency, individual to group learning, emphasis on equal capability over different capability, adapting the school setting to students rather than vice versa, and focusing average students over the best ones (Hernandez-Tutop, 2012).

Standardized Test: An assessment that is administered and scored in a consistent manner and that is typically given to a large population of students (The Glossary of Education Reform, 2015). These can include achievement tests or aptitude/ability tests.

Teacher-based Grade Retention: The practice of not promoting a student to the next grade level, or retaining the student in the current grade, based on school staff recommendation. These recommendations are typically more subjective or anecdotal in nature than test-based

decisions, although the line between teacher-based and test-based grade retention can be blurry (Huddleston, 2014).

Test-based Grade Retention: The practice of not promoting a student to the next grade level, or retaining the student in the current grade, based on student performance on a given assessment. These assessments are typically standards-based and of a minimum-proficiency nature. There is typically a hard pass cut score set on the test, and there may or may not be accompanying policy guidance for making decisions of retention for a student who does not earn a passing score as there is with the IREAD-3 policy in Indiana (IDOE, n.d.).

Summary and Organization of the Study

The use of grade retention as an intervention for students who fail to demonstrate grade-level proficiency has a long history in education; however, it remains a controversial option. This study contributes to the existing body of knowledge on the efficacy of test-based grade retention policies in improving student achievement. It also provides insight into the effectiveness of the IREAD-3 policy in Indiana. In this chapter, a statement of the problem was followed by a statement of the purpose of this study, assumptions and delimitations were articulated, and significant terms were defined for the discussion that follows.

Chapter Two of this study contains a review of related literature including a synopsis of existing research on grade retention, high-stakes testing, and large-scale test-based grade retention policies outside of Indiana. Chapter Three details the methods used for this study, while Chapter Four provides a presentation and analysis of the data obtained. Chapter Five provides the conclusions and implications for further research.

CHAPTER 2

LITERATURE REVIEW

Some people live by the motto "Failure is not an option." For any psychological or pep-talk advantage that may be had from repeating such a credo, the reality in life is that failure is all around us. Contrary to our best wishes, schools are no exception. Students fail every day in different ways and to different extents. When students fail to exhibit evidence of achieving specific minimum academic levels, educators and parents can find themselves between a rock and a hard place in deciding between grade retention and social promotion. Thinking that these are the only two options available for failing students is truly a false dichotomy as there are many other interventions and school reform options that may hold more promise for helping at-risk students (Feuer, 1992; NASP, 2011; Reschly & Christenson, 2013). Regardless, school leaders and education policymakers need to be able to make well-informed decisions in regard to the dilemma of "to retain or not to retain."

School failure has been associated with a whole host of undesirable non-academic outcomes. These include underage alcohol abuse, drug use, depression, higher mortality rates, criminal activity, and need for institutionalization for mental health (Crosnoe, 2006; McCarty, Mason, Kosterman, Hawkins, Lengua, & McCauley; 2008; Richman, Bowen & Woolley, 2004). With this understanding, educators seeking to do anything and everything to help prevent student failure may grasp at whatever intervention they can come up with as a viable solution. Grade retention has a long and varied history in the United States as a persistently found answer to student failure (Hernandez-Tutop, 2012; NASP, 2011). Though grade retention may find its roots in the nineteenth century public school system practice of only assigning promotion to those students whose performance was deemed to merit it, the modern debate regarding the

practice of retention really emerged as a reaction to the practice of social promotion that had become the norm in the schools of the twentieth century (Hernandez-Tutop, 2012) and that has largely been decried as a poor practice.

One reason that grade retention policies have endured through to today is the oft-stated view of many educators that it is bestowing the "gift of time" upon those students who are not successful at their current level (Hwang & Cappella, 2018; Jimerson & Ferguson, 2007; Jimmerson & Renshaw, 2012; Larsen & Akmal, 2007; Renaud, 2013; Raffaele Mendez, Kim, Ferron & Woods, 2015; Silberglitt, Jimerson, Burns, & Appleton, 2006; Stubbs, 2013). While time is a great ally in many undertakings in life and in education in particular, it is certainly not a panacea for all of the many reasons that students experience failure including factors inside and outside of the school. Still, this reasoning, or variations of it, always seems to be the loudest argument for proponents of grade retention.

Method and Organization

In this review of related literature, I searched multiple databases and search engines, primarily OneSearch Ball State University Library, Web of Science, ERIC (EBSCOhost), PsycINFO, and Google Scholar, with terms such as "grade retention," "grade repetition," "social promotion," and "test-based retention." I focused on articles in peer-reviewed journals between the years of 2012 and 2019 before implementing the snowball technique for finding additional sources from the references of the originally located studies. I focused primarily on studies that were quantitative in nature and that analyzed the effect of grade retention on student achievement. In addition to these searches, for purposes of background and related areas of the literature review including conceptual and theoretical frameworks I also used the terms such as "student achievement," "standardized testing," "school accountability," "educational policy,"

"policy mediation," "Rubin causal model," "social ecological model," "ISTEP+," and "IREAD-3" and then applied the snowball technique to find additional references in the studies I had already read.

For the organization and structure of this review, I begin by examining what researchers in the field of education have found about topics that are foundational to the question at the heart of this study: "What is the relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?" This includes a brief history of the practice of grade retention, an examination of high-stakes student achievement testing, and an overview of the effects of grade retention on individual students and their families. Upon this foundational understanding and moving toward matters of policy, I review alternatives to grade retention, such as social promotion, and what research has found regarding the effects of those interventions followed by a look at the academic, non-academic, and economic costs of grade retention. Next, I present what researchers have found regarding the outcomes of grade retention policies by taking a brief look at two separate categories, with a focus on the latter: teacher-based retention policies and test-based retention policies. In conjunction with the examination of test-based retention policies and their application of high-stake achievement assessments, I briefly discuss what research has shown concerning social inequities of such policies, regardless of their scope, whether at the school, district, or state level. I present some of the issues involved in inconsistent implementation of educational policy and its effect on research. Finally, I focus on research that has been done in districts or states with test-based grade retention policies.

Theoretical Framework

As Brofenbrenner (1977, 1979, 1994, 1995, 1999; Bronfenbrenner & Ceci, 1996; Bronfenbrenner & Morris, 1998, 2006) presented the Ecological Systems Theory of Human

Development, there are five levels or spheres of influence in the world of a human being, and the interactions between those levels largely explain the development of the individual. Bronfenbrenner, along with his colleagues, viewed the individual at the center of nested circles that expand outward as follows: (1) microsystem, (2) mesosystem, (3) exosystem, (4) macrosystem, and he later added the (5) chronosystem. An example of the microsystem would be a child's family or teacher/classroom. The mesosystem might be relationships between different microsystems such as that of the parent and teacher. The exosystem might be understood as more removed institutional systems that do not have direct interaction with a student such as the school or school district administration and policies. The macrosystem would be the larger cultural values of the society along with the policies that represent those values as well as state and federal agencies. Finally, the chronosystem represents the differences that can be observed with the passing of time. A visual representation of this model can be seen in Figure 1. Bronfenbrenner and Morris (1998) argued that the environment changes people while at the same time people change the environment.

As his theory evolved over time from the 1970s to the early 2000s before his death, Bronfenbrenner later stressed even more the role that the individual played in their own development, changing the name to Bioecological Systems Theory (Bronfenbrenner, 1994, 1995, 1999; Bronfenbrenner & Ceci, 1996; Bronfenbrenner & Morris, 1998, 2006). In these more mature statements of his theory, Bronfenbrenner put more emphasis on the Process—Person—Context—Time (PPCT) framework as the most appropriate representation of his concepts for use in research. This fit with his concern for the fundamental role of "proximal processes", which are the recurring, bi-directional interactions between an individual and the various levels of their immediate environment (Bronfenbrenner & Morris, 1998, 2006). In fact, in the PPCT

framework, Process plays the most critical role in human development. The second P, Person, considers the characteristics of the individual, including traits such as age, gender, and race as well as mental and emotional resources. The levels of systems (micro, meso, exo, and macro) make up the Context portion. The last element of PPCT is Time, which relates to the passage of time in multiple levels of the context systems in addition to the role of historical events. Tudge, Mokrova, Hatfield, and Karnik (2009) argue that all four elements of the "mature" version of Bronfenbrenner's PPCT should be accounted for in research studies based on this framework to truly evaluate its validity, but they attest that most researchers do not. In fact, many studies only use portions of the earlier versions of Bronfenbrenner's theory (Tudge, Mokrova, Hatfield, & Karnik, 2009).

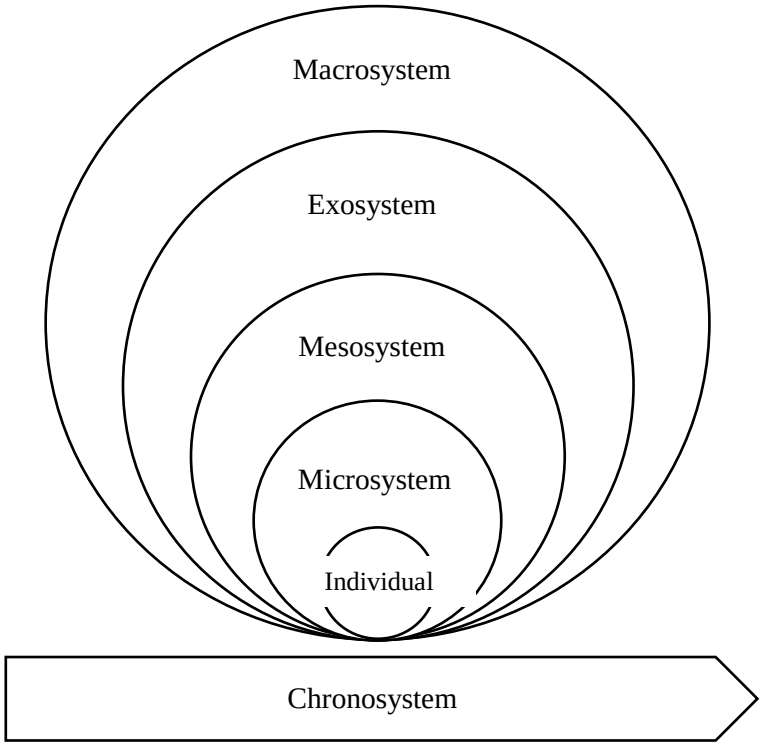


Figure 1. This is a popular representation of the nested systems within Bronfenbrenner's Bioecological Systems Theory. While this can be helpful to visualize the expanding spheres of the context or environment moving out from the individual, some of Bronfenbrenner's descriptions of these systems, such as the mesosytems, can be understood more as networked than as nested.

McLeroy, Bibeau, Steckler, and Glanz (1988) adapted Bronfenbrenner's model to explain behavior of individuals using the following levels: (1) intrapersonal factors, (2) interpersonal processes and primary groups, (3) institutional factors, (4) community factors, and (5) public policy. While this model was presented in the context of health promotion, I believe that it relates well to educational policy. In this study, intrapersonal factors include the basic student demographic data such as age and gender in addition to behavior such as attendance rate. Interpersonal would include family factors such as ethnicity and socioeconomic level. Institutional factors are the school and school district with their specific policies and procedures regarding intervention and grade retention. Public policy is readily seen in Public Law 109 passed in Indiana in 2010 and the subsequent guidance provided by the Indiana Department of Education on IREAD-3 and grade retention. The only one of the five factors identified by McLeroy, Bibeau, Steckler, and Glanz (1988) that is not clearly represented in this study is concept of community factors, there is not a clear differentiation between institutional and community factors in their descriptions. McLeroy, Bibeau, Steckler, and Glanz (1988) instead focus on the various ways that organizations and interactions within Bronfenbrenner's mesosystems affect individual behavior.

History of Grade Retention

Grade retention as an intervention strategy has been implemented in the United States going back to the mid 1800s when the modern system of grade leveling, mostly based on the age of the student, began (Beebe-Frankenberger, Bocian, MacMillan, & Gresham, 2004; Holmes & Matthews, 1984). William H. Maxwell, who became the first superintendent of New York City Schools in 1898, created a report of the district's progress that became the standard for school reporting and included a summary of their grade retention practices which ranged from between

20% and 70% of students not promoted each year (Owings & Kaplan, 2001). C. H. Keyes (1911) studied what he referred to as "acceleration" and "arrests," which is what he called required repetition of a grade, in the New York School system just after the turn of the century. At that time, Keyes stated that these arrests were "present in all schools having a uniform course of study, no matter how free the organization, nor how efficient and numerous the agencies for prevention of arrest" (1911, p. 60). Additionally, Keyes estimated that as many as one-fourth to one-third of all pupils would have to repeat a grade at some point in their academic progress. In more recent decades the proliferation of high-stakes student achievement testing as a part of an educational environment emphasizing standards and accountability has brought about an even greater need for research-based decision-making and policy. Determining the effectiveness of a practice, no matter how widespread, is crucial for policymakers and educational leaders, and so the efficacy of grade retention is the subject of contemporary research.

The modern debate concerning grade retention as a policy can be traced back to the minimum-competency testing that came into being through the late 1970s and early 1980s (Huddleston, 2014; Koretz, 2008). In the face of declining SAT scores (Wirtz, 1977) and growing public perception that educational standards across the country were relaxed in the rise of child-centered curriculum (Berliner & Biddle, 1995; Shepard & Smith, 1989), many called for reform in the public school system that was reportedly failing our nation. The Reagan administration release of *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983) which called for, among other reforms, an increase in testing for accountability purposes provided a rallying point for the angst of the time. A public opinion poll from the mid-1980s revealed that 72% of Americans felt strongly that promotion from grade level to grade level should be based on demonstrated mastery of the

particular requirements of each grade-level (Shepard & Smith, 1989). During the administration of President Bill Clinton, the practice of social promotion came under heavy attack as Clinton repeatedly called for test-based retention policies and touted the existing policy in Chicago as a model for other city districts and states to follow (Huddleston, 2014; Russo, 2005). Policies for grade retention based on failing a minimum reading proficiency test given around grade 3, such as the IREAD-3 policy in Indiana, have gained momentum in the United States since the implementation of this type of policy along with many educational reforms in Florida under the Jeb Bush administration in 2002 (Starr, 2019).

High-Stakes Testing

While high-stakes testing today can be clearly traced back to the practices of schools at the beginning of the twentieth century (Bolon, 2000), the No Child Left Behind Act of 2001 (NCLB) ushered in the current era of accountability and high-stakes testing in the United States. Researchers studied the use of assessments for accountability purposes prior to the current debate, including investigations into the intensely debated and protested use of intelligence test in the 1920s (Bolon, 2000). As Huddleston (2014) argued, this portion of the literature "is relevant to test-based retention policies in that such policies are themselves a form of high-stakes testing" (p. 6). Some researchers have found positive outcomes beyond individual student achievement gains from the implementation of high-stakes testing policies such as improved practices among school staff. Hamilton et al. (2007) found that schools in California, Georgia, and Pennsylvania better aligned their curriculum with state standards and increased their use of student data to make decisions and provide support for struggling students after NCLB tests went into place. Finnigan and Gross (2007) studied teacher motivation at ten elementary schools in Chicago that were placed on probation due to low test scores. They found that while teachers

reported increased effort, openness to trying new approaches, and participation in professional development, they seemed more motivated by how they valued their professional status and by the goals they had for individual students than by the extrinsic threat/incentive placed on them by the policy. Additionally, the longer the schools remained on probation, morale decreased further and negated any motivational affect the policy initially had.

Concerns of the validity of a standardized tests being used as an accurate measurement of student progress have been raised, including those specific to ISTEP+ (Grissmer, Ober, & Beekman, 2014). Studies have also revealed numerous unintended negative outcomes from high-stakes testing policies. Comber (2011) posited that the proliferation of high-stakes, standardized testing can even come to define learning, such as literacy skills. Hout, Elliott, and Freuh (2012) examined how incentives, including the threat of negative consequences, affect student learning in research studies conducted during the NCLB era. They found that while there was a small increase in test scores after such policies went into effect, the gains were explained by score inflation due to teachers "teaching to the test" and narrowing the curriculum. So while there might have been some perceived benefit from implementing the assessments, the gains were not due to the assessments themselves because when they studied similar tests in a low-stakes situation, there was effectively a zero effect. They suggest that high-stakes testing policies do not produce the desired outcomes of the policymakers who put them into place.

Effects on Students and Families

For much of the past several decades, researchers have examined the practice of grade retention and found largely negative results. While proponents of grade retention (e.g., Owen & Ranick, 1977; Winters & Greene, 2006) contend that they are giving students the "gift of time" which is essential for disadvantaged students to achieve mastery of the skills taught in

subsequent years of schooling, opponents (e.g., Huddleston, 2014; Shepard & Smith, 1989) assert that grade retention policies do not deliver on their promise of academic improvement and in fact cause higher dropout rates and disproportionately affect the most at-risk students in our schools. Buckingham, Wheldall, and Beaman-Wheldall (2013) observed that socioeconomically disadvantaged students were more likely to experience confounding issues for achievement such as quality of health, home, and classroom and are more negatively affected by them than their more advantaged peers. Hong and Raudenbush (2005) found that while retention in kindergarten had a negligible effect on overall student achievement scores in school with a retention policy, the negative effects on individual students who were retained were substantial, including losses in both reading and mathematics equivalent to nearly half a year of expected growth.

Moving Toward Policy

When considering options in policymaking, it is important to carefully study the pros and cons of each alternative. In this process of determining the pros and cons of grade retention, a decision-maker may focus on the following areas: (1) whether it is better than alternatives, such as social promotion, (2) whether it is advantageous from an academic perspective, (3) whether it is harmful to students from a psychological perspective, (4) what the associated costs are from an economic perspective, and finally (5) the issues with consistent implementation of large-scale policies. In the next section, I will review some of the findings of previous studies in these areas.

Alternatives to Grade Retention

The first and most obvious alternative to retaining a student in a grade level is social promotion, which is the practice of promoting students on to the next grade level with their age peers regardless of demonstrated mastery of skills. Social promotion has long been viewed as poor practice and was an easy and important target of the modern accountability movement

(Damme, 2016; Denton, 2001; Jimerson, 2001; Lynch, 2014; McMahon, 2018; National Association of School Psychologists, 2011; Vandecandelaere, Vansteelandt, De Fraine, & Van; Fiester, 2013). President Bill Clinton endorsed increased use of standardized testing and grade retention as opposed to social promotion in his State of the Union address three times (Peterson & Hughes, 2011). The research on social promotion as a policy shows that students who are promoted on to the next grade regardless of content mastery are at a greater risk of falling further behind academically than their peers in subsequent years and are more likely to become high school dropouts (Jimerson, Anderson, & Whipple, 2002).

The reality is that it is not an either/or proposition between grade retention and social promotion, and the policy debate should not be framed in this light (National Association of School Psychologists, 2011). Schools can work to identify academically struggling students as early as possible and provide additional supports and remediation. These interventions can occur inside or outside of the regular classroom and within the regular school day and year or as extra programming outside of that in the examples of after-school tutoring or summer school, and such strategies have proven to be effective (Jacob & Lefgren, 2004; Jimerson, 1999). Another option is alternative curricular programming outside the same requirements of state standards and assessments required for a high school diploma, such as what students may receive Special Education services through an Individualized Education Plan (IEP), though this is typically just permissible for schools to do when there is an identified disability. In several studies, researchers have questioned if, rather than the actual retention, it was the additional interventions such as Response-to-Instruction (RtI) tiers and other remediation efforts which account for the initial academic gains in students who are retained in grade level (Denton, 2001; Holmes, 1989).

An increasingly popular early intervention/prevention option for parents in view of higher standards and academic expectations of young students is to "redshirt" their child, which is the practice of delaying the start of schooling by a year (Cannon & Lipscomb, 2011). This is often elected when a child is on the younger side of students who are eligible to begin kindergarten, but can be done in other instances as well depending on the school attendance laws of individual states. Cannon and Lipscomb (2011) found that this practice of redshirting prior to the kindergarten year was correlated with a lower likelihood of later being retained in later grades and. They also found that there were generally positive academic gains for students who were either redshirted or retained in primary grades. Robertson (2021) notes that this practice is often seen as a preferable alternative to grade retention. While Katz (2000) recognizes the growing popularity of this approach and cites several studies that showed positive outcomes, she acknowledges that some studies show that there may be some negative social and behavioral long-term effects for students who were redshirted and that the overall evidence is inconclusive.

Academic Costs of Grade Retention

Academic costs may be the most obvious and sure should be the most important consideration for decisions concerning grade retention or social promotion (Lynch, 2014). While some researchers have found that grade retention produces positive effects on student achievement in the short term after the decision to retain (Gleason, Kwok, & Hughes, 2007), other studies have shown no benefit even for the first year after a student is retained (Burkam, LoGerfo, Ready, & Lee, 2007; Cockx, Picchio, & Baert, 2019). Some of this variance in results can be attributed to different comparison models. For example, whether the students studied are compared to their grade-level peers or age-level peers. Regardless of any short-term gains, studies have also found a correlation between grade retention and several negative academic

effects including lower participation in higher level courses, increased dropout rates, and lower enrollment rates in post-secondary schools (Cockx, Picchio, & Baert, 2019; Fine & Davis, 2003; Hong & Raudenbush, 2005; Ou & Reynolds, 2008; Stearns, Moller, Blau, & Potochnick, 2007).

Non-academic Costs of Grade Retention

Beyond a consideration of social-emotional outcomes for students, some studies have shown that the effects of grade retention may well extend beyond the individual students who are retained and their families. Fanguy and Mathis (2012) found that among the many negative psychosocial effects of grade retention on students was a resentment of teachers and school administrators. This may present in multiple ways including apathy toward school or heightened emotional outbursts which would require additional time and attention from school staff. Holmes and Matthews (1984) concluded from their meta-analysis that students who were retained were not as personally well-adjusted and had more negative attitudes about school than their promoted peers. This is in addition to the link studies have found between grade retention and increased behavior issues and suspension rates (Fanguy & Mathis, 2012). Some of the connections between grade retention and negative effects can raise a "causation versus correlation" debate and raises additional questions, as evidenced by the link between Adverse Childhood Experiences (ACEs) and grade retention (Hinojosa, Hinojosa, Bright, & Nguyen, 2019).

Economic Costs of Grade Retention

The number of students involved in policies such as the IREAD-3 retention policy is substantial. There are not readily available data on retention rates in Indiana according to the IREAD-3 retention policy. Roach and Kloosterman (2014) also referenced the lack of published data on the number of students retained in a given academic year by the Indiana Department of

Education, and they used the published numbers of the pass rate and the Good Cause Exemption rate to find the maximum percentage of students who would have been retained. This does require an assumption that all school leaders are following a strict interpretation of the IREAD-3 policy. However, schools do have leeway in the IREAD-3 guidance sent out by the Indiana Department of Education to be creative with schedules and move a failing student on to fourth grade with the assurance that they will also continue to receive instruction in third grade reading.

Since there is not a great deal of oversight concerning the fidelity with which schools follow the guidance, it is impossible to know for sure how many students are affected by this state-level policy. To approximate how many students are affected by the IREAD-3 policy, I took a similar approach to what Roach and Kloosterman (2014) used for their analysis. By looking at the released statewide data on the Indiana Department of Education website for IREAD-3 and subtracting both the pass rate and the Good Cause Exemption rate from 100%, I got the potential retention rate for each year since 2012. This value would be the highest number of students who would have been affected by the IREAD-3 retention policy, but it does not include any students who might have been retained for other reasons determined locally apart from the IREAD-3 policy. Figure 2 shows the released pass rate data along with Good Cause Exemptions for IREAD-3 over the years from 2012-2019. (There was no administration of the spring 2020 IREAD-3 assessment due to the COVID-19 pandemic.) Those numbers put an annual average of potentially retained 3rd graders at 5.2%, which means that more than 4,000 students could be retained each year. Using the figure from the U.S. Census Bureau of $10,045 spent per pupil in Indiana in 2017, retaining 4,000 students would cost $40,180,000.

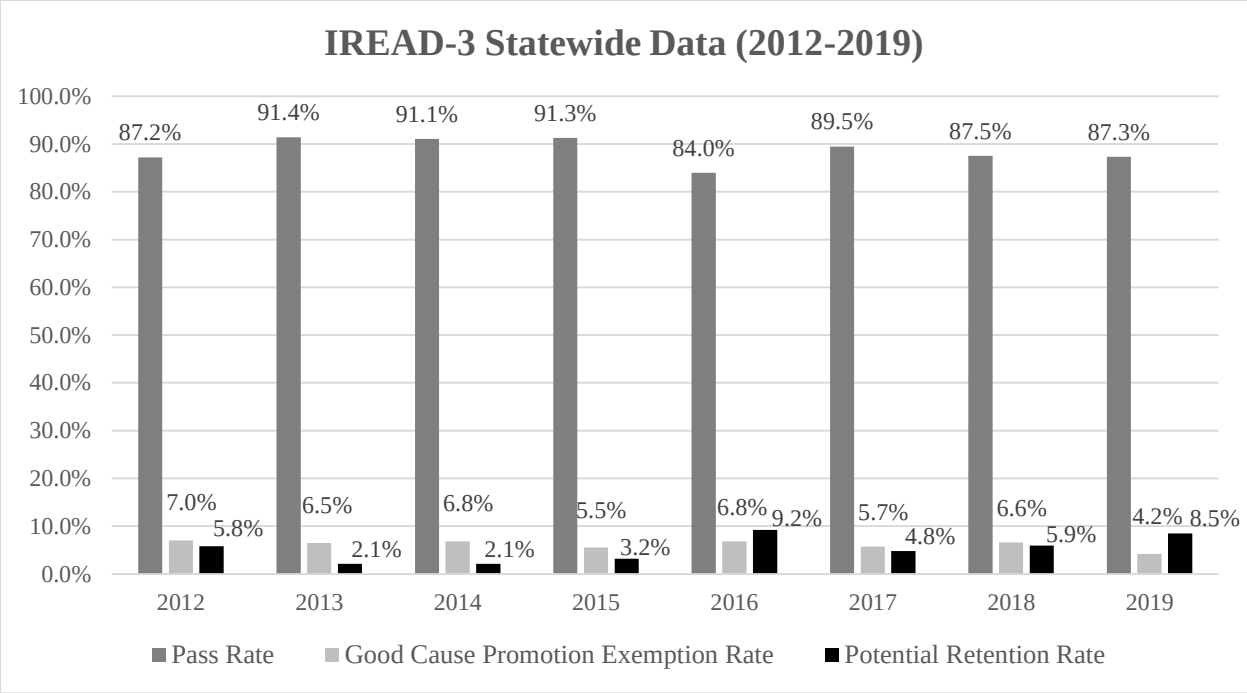


Figure 2. The statewide IREAD-3 data from all 3rd graders in Indiana for the years from 2012 (when IREAD-3 was first given) through 2019. The figure presents the overall percentage of students who passed, the percentage of students who received a Good Cause Exemption, and the remaining percentage of students who were presumably retained according to the IREAD-3 grade retention policy. Data retrieved from Compass on IDOE website.

Many researchers have found that grade retention leads to higher dropout rates or, in other words, a lower likelihood of graduating from high school (Eren, Lovenheim, & Mocan, 2018; Jimerson, 2001; Holmes, 1989), although in some cases the effect on graduation rates was dependent on the age of the student when they were retained (Manacorda, 2012; Jacob & Lefgren, 2009). Additionally, the effects of grade retention and associated costs may extend far beyond high school. In studying the effects of a test-based grade retention policy for 8th graders in Louisiana, Eren, Lovenheim, and Mocan (2018) found that there was a positive relationship between grade retention and involvement in criminal activity by the age of 25. The largest observed effects were in violent crime. However, since Eren, Lovenheim, and Mocan (2018) did not find a related increase in juvenile crime, it is likely that the detected effect was more from increased likelihood of not graduating from high school and being less invested in their

educational opportunities along with the associated employment and economic effects that brings. Regarding the financial impact related to crime of the test-based retention policy in Louisiana, they estimate a range between $2.6 and $18.4 million.

Teacher-based Grade Retention

Teacher-based grade retention is the type of retention that has been studied far more frequently in the academic literature, and, over the past 60 years, "has produced some of the most consistent findings in the research literature" (Huddleston, 2014, p. 8). In addition to original studies that have been performed, there are abundant examples of meta-analyses (e.g., Holmes, 1989; Holmes & Matthews, 1984; Jimerson, 2001) and literature reviews (e.g., Huddleston, 2014; Jimerson, Anderson, & Whipple, 2002; Shepard & Smith, 1989; Valbuena, Mediavilla, Choi, & Gil, 2020; Xia & Kirby, 2009). Again, these have largely found no effect or negative outcomes from the practice of grade retention. However, in consideration of the validity of the practice, it is important to understand that many studies on grade retention have not made a distinction between teacher-based and test-based decisions, which somewhat confuses the issue. Additionally, there has been a great deal of variability in sample sizes and the rigor of methodology (Xia & Kirby, 2009).

Holmes (1989) found a positive effect from grade retention in only nine of the sixty-four studies in his meta-analysis. Each of the nine with positive outcomes also included other interventions such as early identification and individualized learning supports and had an unusually high number of White and middle-class students with better than average IQ scores. Additionally, none of these nine studies tracked student achievement effects beyond the first year following retention. Nevertheless, some studies have shown that grade retention can produce positive effects or at least prevent future negative effects. Hong and Yu (2008) performed a

rigorous analysis of data on kindergarten retention and found no harm as a result of the retention and that it would have been more likely for the retained students to have experienced issues with self-confidence and behavior had they been socially promoted. These findings are mitigated by even the same researchers themselves who found there was not an overall benefit to retention in primary grades as positive or negative effects fade over time (Hong & Yu, 2007).

Test-based Grade Retention

According to Huddleston (2014), "the findings on test-based retention are mixed and based on a more limited pool of studies" (p. 12). Inconsistency in the implementation of test-based retention policies may explain some of the variation in outcomes that have been observed in the various studies (Greene & Winters, 2007). Some researchers have observed the different outcomes of research studies and determined that issues with test scores across different grade levels, the composition of comparison groups of students that were either promoted or retained, and the points where effects on achievement are measured account for the variations (Roderick & Nagaoka, 2005).

The effects of test-based retention policies reach beyond the students who are actually retained (Greene & Winters, 2007; Winter, 2019). Schools institute supplementary supports such as tutoring outside of the regular day and over the summer in addition to running remedial support groups and courses during regular hours. Beyond the retained students, some studies have suggested that the test-based retention policies have an effect on the academic achievement of non-retained classmates (Allensworth & Nagaoka, 2010; Greene & Winters, 2007)--mostly as it could be a motivating factor for students or provide greater quality or quantity of instructional opportunities as the schools respond to the looming threat of retention, but also as these policies can change the make-up of the peer groups.

Social Inequities

A persistent issue in our nation's schools during the drive for higher levels of learning and achievement for all students has been finding ways to simultaneously close the achievement gap (Goddard, Skrla, & Salloum, 2017). Unsurprisingly, there have been concerns raised over social inequalities of grade retention policies (Huddleston, 2014). Ethnicity, gender, socio-economic status, disability status, and parental characteristics all appear to be related to predicting rates of retention (Adams, Robelen, & Shah, 2012; Frey, 2005; Hughes, Cao, West, Smith, & Cerda, 2017; Skiba & Rausch, 2004; Yang, 2019). As an example, the average retention rates for students of different ethnicities from 1994-2016 have been 3.8% for Blacks, 3.2% for Hispanics, and 2.1% for Whites, taken from Status and Trends in the Education of Racial and Ethnic Groups from National Center for Educational Statistics (NCES). There is also a clear difference in retention rates between genders as the average retention rates from 1994-2016 were 2.8% of males and 2.4% of females (NCES, 2017). Booher-Jennings (2008) found that there were differences in the way that teachers perceived the academic failures of boys and girls in her study of the retention policy in Texas. In her study, she noted that teachers tended to tell girls that doing their best was enough, despite any failure experienced on a high-stakes test. Contrariwise, the failing boys were rebuked for failure to behave or listen properly and for their lack of effort. Further, Booher-Jennings (2008) observed that boys for whom the cause of failure was misdiagnosed began to lose faith in the value of putting forth effort in the academic setting, thus initiating a self-fulfilling prophecy of failure.

Inequities in retention policies between ethnicities, socio-economic levels, and gender can be seen in varying school types across the nation (Hwang & Cappella, 2018). Interestingly, Hong and Raudenbush (2005) found that schools from across the United States with a retention

policy ostensibly had many advantages over schools that did not have a retention policy, from data collected in 1998-1999. The retention schools were on average more likely to be nonpublic, suburban, lower in minority diversity, smaller class sizes, higher levels of parent involvement, and higher levels of order. It was not a part of their study to investigate the reasons for these differences, but the differences were remarkably consistent. Interestingly, a study of high school students in Belgium performed by Cockx, Picchio, and Baert (2019) found that the portion of the student population most negatively affected by being retained were the students with lower ability. This conclusion was separate from the fact that students with lower academic abilities are more likely to be considered for retention. This conclusion is notable and somewhat counterintuitive as it may generally be believed that the lowest students are the only ones who should be considered for grade retention.

Inconsistencies in Implementation

While universal accountability policies such as test-based grade retention are a favored instrument of policymakers to attempt to address inequity, blindly trusting that the implementation will be even and consistent can be a crucial miscalculation. It is equally problematic to assume that individual districts, schools, and educators simply choose to adhere to policy or ignore it (Spillane, 2005). Spillane describes sensemaking based on existing values that occurs at each level of implementation of a policy. This makes the task of any kind of school reform that much more difficult. In a similar discussion of policy in education and its implementation, Ball, Maguire, and Braun (2012) define the idea of "policy" as follows:

> Policy is complexly encoded in texts and artefacts and it is decoded (and recoded) in equally complex ways. To talk of decoding and recoding suggests that policy 'making' is a process of understanding and translating – which of course it is. Nonetheless, policy

> making, or rather, enactment is far more subtle and sometimes inchoate than the neat binary of decoding and recoding indicates. (p. 3)

They further discuss the multitude of factors that come into play for how any outside policy will be enacted within the context of a specific school. This is similar to the existing values concept that Spillane (2005) referenced.

As one example of the variation that can occur in districts across a state due to various factors, LiCalsi, Ozek, and Figlio (2019) found that Florida's third grade reading skills retention policy showed evidence of variability in enforcement based on socioeconomic background of the student's family. This difference was particularly evident for students whose mothers had higher levels of education, possibly due to the fact that they were better at negotiating exceptions to the retention policy.

Large-scale Policies (Citywide and Statewide)

Much of the research on test-based retention policies has been performed in cities or states where such policies have existed for several years. These specific geographic areas include the cities of Chicago and New York City as well as the states of Florida, Texas, Georgia, Wisconsin, and Louisiana (Huddleston, 2014). Chicago was the earliest to implement a test-based grade retention policy in 1996, while the others followed with their implementation occurring between 2000 and 2004. They all included either third or fourth grade as the initial year of a grade retention action due to failure to achieve a proficiency score on an assessment, and many of them include two or three other grade levels as "gates" where a minimum score is required. At the time of this writing, each of these policies remain in effect with the exception of Louisiana, which repealed the original law in 2009 (Huddleston, 2014). While each of these policies share many similarities in their structure and requirements, they are also a few areas of

difference. A significant difference is that some of the policies require a passing score on a single assessment at designated gateway grades (Texas, Georgia, Louisiana, and Chicago) while others allow for additional information such as a portfolio or alternate standardized tests to be considered in the retention decision (Huddleston, 2014). They also have widely varying rates of retention, anywhere from 12-14% in Florida (Winters & Greene, 2012) to 1% in New York City (McCombs et al., 2009). According to Mordica (2006) and Henry, Rickman, Fortner, and Henrick, (2005), 61-68% of third graders who would have been retained due to failing the Criterion-Referenced Competency Tests in the 2003-2004 school year were still placed into fourth grade through the appeals process.

The timing of the grade retention seems to be significant as well. Jacob and Lefgren (2009), in analyzing data from Chicago's test-based grade retention policy, found that students retained in third grade experienced a small gain in achievement while sixth graders did not. They also found that sixth graders who were retained saw no negative effect on their graduation rates while students retained in eighth grade had a lower rate of completing their high school diplomas. Similarly, Manacorda (2012) found that students retained in junior high in Uruguay increased the rate of dropping out before completing high school.

It is often believed that grade retention in primary grades, kindergarten through second grade, is the most effective and least harmful to students, although others consider retention in kindergarten to be especially inappropriate (Hong & Raudenbush, 2005). From that viewpoint, Hong and Raudenbush (2005) researched the effect of grade retention in kindergarten using data from the US Early Childhood Longitudinal Study Kindergarten (ECLS-K) cohort. This dataset comes from a longitudinal study of the kindergarten class of 1998-99 through their eighth grade year. It contains information about the students, their families, and their schools gathered

through student assessments along with interviews and questionnaires of parents, teachers, administrators, and the students themselves. Hong and Raudenbush used 11,843 students from schools across the United States in their analyses. In addition to examining the effect on individual students who were retained, they investigated whether a retention policy in kindergarten increased achievement for the entire school due to providing more homogenous levels of academic achievement within classrooms and thus making it easier to teach those groups at a more ideal level and pace.

Hong and Raudenbush (2005) produced some interesting findings from their study. They found that the average effect of a school implementing a policy of retention was negligible for students who met the criteria of being at-risk for retention based on their prior academic performance. The effect was also negligible for the other students who were promoted in schools with retention policies, which is contrary to the belief of some policymakers that those students should benefit from a more homogenous class in the following year after students were retained. They found that students who were retained scored lower, in both reading and mathematics, than their statistical analysis would have predicted for them had they been promoted. This effect was somewhat dependent on how greatly in need of repeating a grade a student was. For example, the students who were identified as most at-risk for repeating were the only students that did not seem to perform lower than their predicted potential. In summary, Hong and Raudenbush found that there is no statistical support, at a point one academic year after the retention, for a policy of retention in kindergarten as there is not a positive effect from such a policy for the average assessment scores of the school, for individual students who are promoted on to the next grade, or for individual students who are retained.

Lorence (2014) studied data from the years immediately prior to the implementation of a test-based grade retention policy in Texas that focused on third grade reading proficiency. He found not only did students who were retained in third grade due to low scores on a reading assessment outperform students who had been socially promoted in the year immediately following the retention, but they also continued to demonstrate positive effects from the retention in subsequent years (effect sizes of 1.17, 0.98, 0.80, 0.59, 0.66, 0.38 for each year from third through eighth grade). Notably, neither the students retained nor the ones who were socially promoted were able to achieve at an average level compared to reading scores of their peers. The positive effects Lorence observed in Texas may be at least partially explained by the additional interventions put in place for retained students and not just the fact of repeating third grade.

The Florida policy, unlike the test-based retention policies in Chicago and New York City which have thresholds for both mathematics and reading, focuses solely on a reading proficiency test, similar to the Indiana policy. Just as the Indiana policy includes "good cause exemptions," the Florida policy allows for exemptions for students with IEPs that specify that the state assessment is not appropriate or who have been retained previously in third grade, Limited English Proficiency (LEP) students with less than two years of instruction in English, any student who had already been retained twice, students who score above the 51st percentile on a different nationally normed reading test, or students who demonstrate reading proficiency via a portfolio of their work (Schwerdt, West, & Winters, 2017). Even with all of these exemption options, Florida saw a sharp increase in retention rates for third graders in 2003, the first year of implementing the test-based retention policy—13.5% up from 2.8% the prior year before steadily declining over the next five years to 5.6% in 2008 (Schwerdt, West, & Winters, 2017).

A study of particular note relating to my study on the effects of the IREAD-3 grade retention policy in Indiana is the study of the Florida reading test-based grade retention policy by Schwerdt, West, and Winters (2017). They examined the effect the policy has had on student outcomes over time. This was different from other studies (e.g. Greene & Winters, 2007; Jacob & Lefgren, 2009) that focused on student outcomes within the first two years following the grade retention and thus could not analyze whether there was a wearing off effect over time. Notably, the Florida policy requires schools to implement additional interventions including the option of summer instruction in reading, assignment in a "high-performing" teacher's classroom, and intensive interventions throughout the school year (Schwerdt, West, & Winters, 2017). As a result, this study measured the collective effect of grade retention with the additional interventions, and this differs from some other studies on retention. Schwerdt, West, and Winters (2017) found that retained students under the Florida policy experienced short-term gains, and, interestingly, those gains came in both reading (+0.31 standard deviations) and mathematics achievement (+0.23 standard deviations). They also found that the achievement gains in both reading and mathematics did fade over time, to a point of being no longer statistically significant within five years. They also found that students being retained in third grade by this test-based policy led to higher grade-point averages, fewer remedial courses, and no effect on the student's likelihood of graduating from high school, though it did delay graduation by 0.63 years. This study supports the findings of Winters and Greene (2012), whose research shows that students retained in third grade in Florida saw gains in achievement that lasted through the eighth grade.

It can be questioned whether any gains seen on a large-scale level are a result of the actual retention as an intervention or from the effect that such policies have on educational

leaders and the decisions they make, including the provision of supplemental interventions (Duke, Moje, & Palincsar, 2014). In a meta-analysis, Robinson, Lloyd, and Rowe (2008) found that educational leaders had a significant effect on student outcomes, especially as it relates to areas such as goal setting, resource allocation, and teacher development.

Summary

The body of extant research discussed above seems to provide the basis for several conclusions. With decades of research on the negative effects of grade retention in view, it may be appropriate to view the burden of proof to be on proponents of grade retention to demonstrate that their retention policies produce successful outcomes (Holmes & Matthews, 1984). However, the body of research on the effects of grade retention is mixed, particularly when more recent studies concerning test-based retention are considered.

While high-stakes testing does not produce the desired outcomes of the policymakers who continue to turn to them for education reform efforts, utilizing these assessments to inform decisions about grade retention (i.e. test-based retention) appears to be much more likely to produce beneficial outcomes for students than simply retaining a student based on the fact that the teacher observed a lack of engagement or pattern of failure at a given grade level (i.e. teacher-based grade retention). The questions remain of how effective these policies are and whether they should be an accepted practice. Concerns remain about negative effects on students, families, classmates, and teachers along with fears about the social inequities inherent in the policies, and this does not even account for the mixed results concerning the effectiveness of test-based retention on the most direct expected positive outcome—increased academic achievement.

In the next chapter, I will focus on the research methods, which include the design of my research study and descriptive statistics of the dataset. After that, I will outline the independent and dependent variables and my plan for analysis. I will conclude the chapter with the limitations of this study.

CHAPTER 3

RESEARCH METHODS

This chapter begins with a description of the purpose of the study and enumeration of the research questions. After that, I discuss the research design and the sample used. This includes a brief description of how the archival data were obtained, the independent and dependent variables used in the study, and a variety of descriptive statistics on the dataset. I then conclude with a discussion of the limitations of this study and a summary of the chapter.

The purpose of this study is to analyze the effectiveness of test-based grade retention as a policy and, specifically, to analyze the effectiveness of the IREAD-3 policy in the state of Indiana. The focus of this analysis is on the academic achievement of the student rather than psychological or social effects. Specifically, student testing data from the Indiana Department of Education will be used to examine the effects of grade retention in third grade in accordance with the IREAD-3 policy on the achievement levels of those students on the annually administered Indiana Statewide Testing for Educational Progress Plus (ISTEP+) in the years after the decision to retain the student in grade. Test data from Grades 3-8 are used in analysis. The independent variables are the IREAD-3 pass status and the decision to retain or not. The dependent variables are ISTEP+ scores in Grades 4-8.

Purpose of the Study

The purpose of this study is to analyze the effectiveness of test-based grade retention as a policy and, specifically, to analyze the effectiveness of the IREAD-3 policy in the state of Indiana. The focus of this analysis is on students' academic achievement rather than psychological or social effects. Student testing data from the Indiana Department of Education will be used to examine the effects of grade retention in third grade in accordance with the

IREAD-3 policy on the achievement levels of those students on the annually administered Indiana Statewide Testing for Educational Progress Plus (ISTEP+) in the years after the decision to retain the student in grade. Test data from Grades 3-8 will be used in analysis. The independent variables are the IREAD-3 score and the decision to retain or not. The dependent variables are ISTEP+ scores in Grades 4-8. The following variables will be controlled for as they have been shown to be associated with student achievement levels as measured by standardized tests: free/reduced lunch status (a measure of socioeconomic status), IEP (Individualized Education Plan) / disability status, ELL (English Language Learner) status, ethnicity, and gender.

Research Questions

While this study contributes to an understanding of the effectiveness of test-based retention policies in general, the specific research questions of this study relate to effects on student academic achievement of the IREAD-3 policy in the state of Indiana. This study has one primary research question and two ancillary questions that will directly inform the primary question.

Primary Research Question: What is the relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?

Subquestion 1: What is the relationship between grade retention and student achievement at each grade level 4-8 after the decision to retain?

Subquestion 2: What is the relationship between grade retention and student achievement considered longitudinally across grade levels 4-8 after the decision to retain?

The choice to use grades 4-8 after the decision to retain arises from the current grade level where retention is being considered (Grade 3) and the data available. This range of grades is also consistent with similar studies. Significantly for the purpose of using statewide data in Indiana, after the decision to retain in Grade 3 there is continuous data from the ISTEP+ assessment through Grade 8 over the years of 2012-2018. This also follows the students through to the beginning of high school.

Research Design

Though researchers have studied grade retention for more than a century (see for example, Keyes, 1911), there is a lack of consistent findings. A major reason for the inconsistencies that have existed in conclusions has been issues in design and methodology (Beebe-Frankenberger, Bocian, MacMillan, & Gresham, 2004; Frey, 2005; Hong & Raudenbush, 2005; Huddleston, 2014; Lorence, 2014; Wu, West, & Hughes, 2008). I have chosen to use a quasi-experimental quantitative approach to study the effects of grade retention due to the IREAD-3 policy on student achievement levels. As Creswell (2009) discusses, quantitative research studies get at the relationships among numerically measured variables and is often used when seeking to understand the "why" of differences between two or more groups. This is primarily a longitudinal study for students in the first cohort to take IREAD-3 under the policy in 2012 as they progressed from third through eighth grade and is conducted using a Control-Group Interrupted Time-Series Design (Creswell, 2009).

I focus on a comparison of students who did not pass the IREAD-3 assessment who were then retained or promoted. I want to examine whether the students were, in fact, retained in Grade 3 according to the policy and explore how they performed on subsequent ISTEP+ tests through Grade 8 in both English/Language Arts and Mathematics. While this study contributes

to an understanding of the effectiveness of test-based retention policies in general, the specific research questions of this study relate to effects on student academic achievement of the IREAD-3 policy in the state of Indiana.

Data Collection

I submitted a request for data release to the Indiana Department of Education in November 2019 (the DSA Proposal can be found in Appendix A). The research questions listed on the DSA Proposal differ slightly from the final versions used in the completion of this dissertation. A committee reviewed my request, and I was contacted to set up a conference call on December 2, 2019 to discuss my proposal. During that call the participants approved my request and informed me they would send it on for sign-off by the state superintendent. In March 2020 I received a signed DSA (Data Share Agreement) from the IDOE which I also signed and returned. Due to issues likely related to the COVID-19 response, I found out in early April 2020 that I had to resubmit the signed DSA. I received access to download the data files on April 16, 2020. There was a single data file for each year from 2012-2018. I was able to use these files to assemble a comprehensive dataset for those years with a focus on just the 2012 Grade 3 cohort.

Description of the Dataset

I secured permission for data released from the Indiana Department of Education for all third graders who took the Indiana Reading Evaluation and Determination (IREAD-3) assessment in the spring of 2012. As a part of this release, I obtained student achievement data in English/Language Arts and Mathematics for these students in subsequent years as measured by the Indiana Statewide Testing for Educational Progress-Plus (ISTEP+) assessments given

each spring from 2012-2018. Table 1 provides an overview of the proportion of different demographic factors in this dataset.

It is not clear why there would be so many "Unknowns" in the pay status categories. This oddity was the reason for exclusion of many student records in later analysis. Also, the pass rates, retention rate, and other percentages are slightly different in this dataset compared to the released figures from the state. This could be due to specific qualifications of inclusion in the publicly released statistics or variability in collection/reporting from schools. For example, there are times when the state does not include students with indetermined scores and other times when they are counted along with the Did Not Pass group. The differences between what I calculated from my dataset and the publicly released figures have no effect on the validity of this study in answering the research questions, but I wanted to point out the small discrepancies in some of the pass rate information listed in Table 2 compared to those publicly released values.

Pursuant to my data request submitted to the Indiana Department of Education, I received one data file for each year from 2012 to 2018. In each file I was able to access the following variables for each case (explanatory information is included for some variables with specific categories):

1. School Corporation ID
2. School Corporation Name
3. School ID
4. School Name
5. Student Alternate ID
6. Grade
7. Days Attended (Indiana requires a minimum of 180 days in each school year)
8. Birth Date
9. Gender (This dataset includes the following categories: Female and Male)
10. Ethnicity (This dataset includes the following categories: American Indian, Asian, Black, Hispanic, Multiracial, Native Hawaiian or Other Pacific Islander, and White)
11. Free/Reduced Lunch Status (This dataset includes the following categories: Free/Reduced price meals, Paid meals, and Unknown)
12. IREAD IEP Exemption (yes/no)

13. IREAD ELL Exemption (yes/no)
14. IREAD Score
15. ISTEP+ Math Proficiency (This dataset includes the following categories: Did Not Pass, Pass, and Pass +)
16. ISTEP+ ELA Proficiency (This dataset includes the following categories: Did Not Pass, Pass, and Pass +)
17. ISTEP+ Math Scale Score
18. ISTEP+ ELA Scale Score

Table 1

Demographics Breakdown of 2012 Grade 3 Cohort

Student Category	% of population
Female	48.5%
Male	51.5%
American Indian	0.2%
Asian	2.0%
Black	11.8%
Hispanic	10.4%
Multiracial	4.7%
Native Hawaiian or Other Pacific Islander	0.0%
White	70.9%
Free / Reduced meals	48.3%
Paid meals	44.3%
Unknown pay status	7.5%

Table 2

IREAD-3 Pass Rates by Demographic Category for 2012 Grade 3 Cohort

Demographic Category	IREAD-3 Pass #	IREAD-3 Test #	IREAD-3 Pass %
Female	38638	41493	93.1%
Male	39518	44018	89.8%
American Indian	191	202	94.6%
Asian	1470	1684	87.3%
Black	8180	10102	81.0%
Hispanic	7513	8853	84.9%
Multiracial	3611	3986	90.6%
Native Hawaiian or Other Pacific Islander	40	42	95.2%
White	57151	60642	94.2%
Free / Reduced meals	35634	41262	86.4%
Paid meals	36487	37854	96.4%
Unknown pay status	6035	6395	94.4%

To begin working with these datasets released from the Indiana Department of Education, I renamed several variables, calculated several more, and recoded other variables before merging the various years into a single data file. I began by renaming variables to include the year in their name. I summed, using the aggregate function of SPSS, the days attended in multiple schools by a single student in a single year as the Sum of Days Attended to attempt to get a "total

days attended" value for each student. I then identified duplicates and sorted on days attended in the multiple schools for students who experienced mobility and removed all but the most attended school so that each student would have only one case each year. This resulted in the "Days Attended" for each year being the highest number of days a student attended at a single school in that year, and it also meant that the only reported school in a given year was the one attended for the most days. An area where many anomalies were evident in the dataset was the reported days attended. There is no way of knowing how many days the individual schools were actually in session, but the vast majority of public schools in Indiana follow the minimum requirement of 180 full days in each academic year -- there are some schools that go additional days, and it is possible that there could be some that had less than 180. Table 3 provides overall information about the variables of Days Attended in 2012 and Sum of Days Attended in 2012, which I described above. Table 3 displays that the sum of days attended has less variability but also brings in a few data reporting discrepancies which result in a maximum figure such as

Table 3

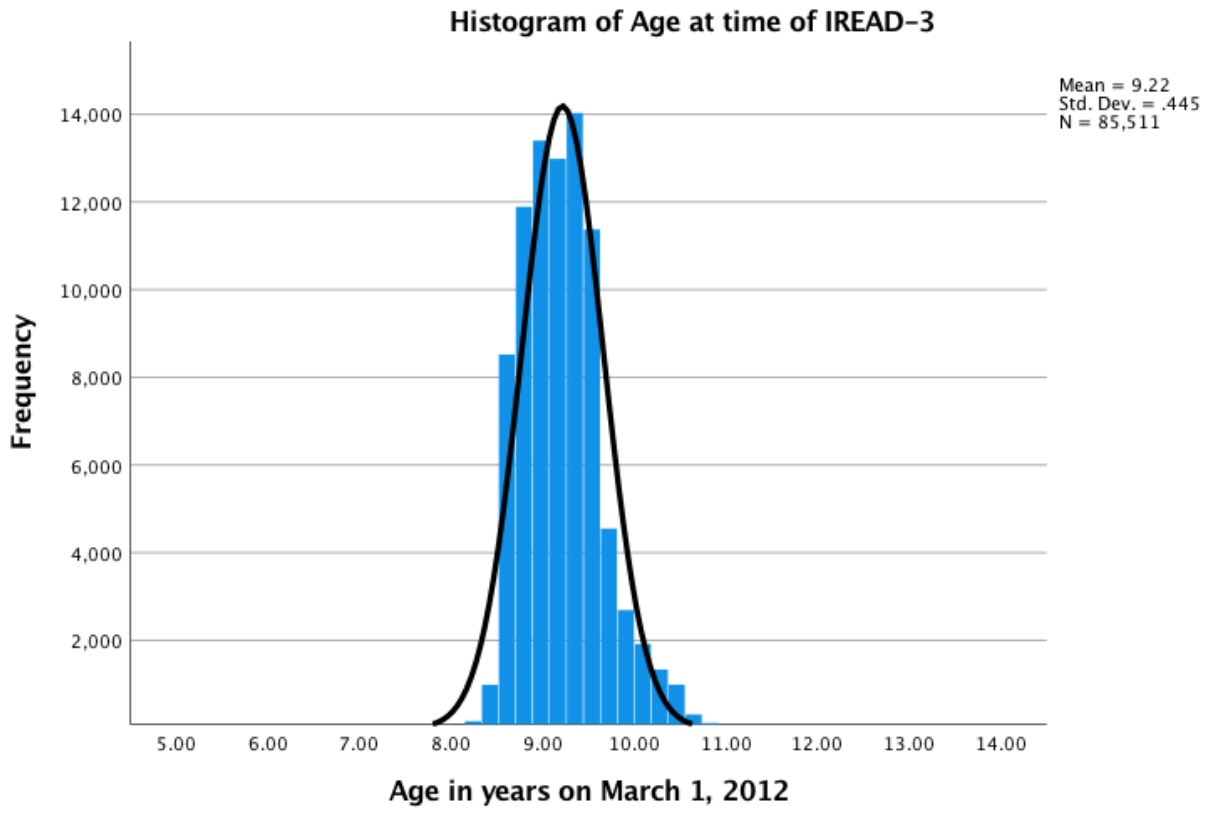
Descriptive Statistics of Attendance Variables for 2012

	Days Attended (at a single school)	Sum of Days Attended
N	85511	85511
Mean	167.6	171.1
Median	175.5	176.0
S. D.	25.86	23.72
Minimum	1.0	1.0
Maximum	210.0	367.0

attending school for 367 days in a year. Using the sum makes more sense in trying to relate the number of days attended in a school year to student performance in later calculations as this variable is only intended to account for the proportion of the entire year the student attended school. Mobility between schools is considered in a different factor.

As a next step, I standardized the test scores for each year using all reported test values from that year. This meant that the z-scores for the Mathematics and English / Language Arts assessments were calculated on values from Grades 3-8 in each year. Since it is for comparison's sake rather than being tied to a specific test score value, standardizing the test scores was a better choice than trying to make sense of scale scores and accounting for the change in the ISTEP+ assessment itself that occurred in 2015.

Figure 3



Since differences in age could indicate many things, including retentions prior to Grade 3, I calculated age for each student as of March 1, 2013. I chose that date since the IREAD-3 spring test window was in March. For purposes of statistical analysis, the specific date chosen is not as important as the ability to compare relative ages. As in many cases of a dataset of this size, some individual variable values were odd. The age based on birthdate had some anomalies in the reported data. The minimum age reported was 5.55 years and the maximum was 22 years. Since there is no way to explain the anomalies in the dataset, I also could not appropriately verify their validity. A histogram of the ages of the 2012 Grade 3 Cohort is provided in Figure 3.

Data Analysis

The following variables are controlled for in the longitudinal multilevel analysis as they have been shown to have an association with student achievement levels as measured by standardized tests: IEP exemption status (for identified disabilities), ELL exemption status (for English Language Learners), school attendance rate, Free/Reduced lunch status (a proxy for socioeconomic level), ethnicity, and student mobility (as measured by moving between schools during the same school year). I also controlled for gender and the age of the student when they took IREAD-3 in 2012. In addition to these, I included the student's school district, known as the school corporation in Indiana, due to the variance in how the IREAD-3 retention policy is implemented locally, including levels of additional interventions provided and likelihood of actually retaining according to the policy.

There is disagreement among researchers on whether to use same-grade or same-age comparisons in studies on the effects of grade retention, and these two different approaches may account for some of the disparities in various studies (Lorence, 2014). In the first option, same-grade, comparisons of measurements are made of retained students to their same-grade peers

meaning there is a lag of one year in the comparisons. For example, a student who was retained would have their 6th grade achievement scores from 2015 compared to their original cohort peers' 6th grade scores from 2014. In same-age, the comparisons of test scores are made between students in the same year who are the same age. For example, a student who was retained would have their 5th grade achievement scores from 2015 compared to the 6th grade scores or their original cohort peers' 6th grade scores from 2015. While I believe that same-grade is generally a better comparison point when considering the effectiveness of grade retention, since I will be using results from ISTEP+, which is a criterion-referenced test based on specific grade levels, the same-grade approach is the only way to appropriately analyze such data. In other words, there is not a truly aligned scale score from the ISTEP+ data, making any grade-to-grade comparisons extremely problematic.

In the earlier parts of analysis, I use t-tests on the means of subgroups prior to utilizing longitudinal multilevel modeling as a part of my data analysis to examine the effects of retention on academic achievement in subsequent years. All analyses were completed separately for both English/Language Arts and Mathematics sections of the ISTEP+. I decided to use standardized scale scores at each grade level to account for different ranges of test scores at each grade level and to mitigate the effects of the change in the ISTEP+ assessment itself that occurred in 2015. This use of z-scores is somewhat similar to what Grissmer, Ober, and Beekman (2014) applied in their analysis of ISTEP+ data over multiple years. In the final part of my data analysis, I employ a longitudinal multilevel modeling approach that controls for numerous covariates, including student demographics and achievement prior to retention as recommended as a higher design quality by Allen, Chen, Willson, and Hughes (2009). In addition to these factors, the school

district was included in a level of the modeling in order to account for variability in the differences in local policies regarding retention and remediation.

Limitations

While using data from the Indiana Department of Education allowed for a very large sample size, the data and variables available limit the study in controlling for additional possible variables that affect student achievement. This study is also entirely dependent on the accuracy of the reported information. Additionally, the available dataset relied on ISTEP+ assessment data as the sole measure of student achievement, which is a narrow perspective at best. The change in the ISTEP+ test in 2015, which occurs in the middle of the range of years including in this study, is another potentially limiting factor to the reliability of comparisons made over the entire span of years from 2012-2018. However, I attempted to mitigate that issue through the use of standardized scores.

Another significant limitation of this study is the issue of policy mediation. Different districts and even individual schools vary greatly in available remediation supports and interventions beyond the regular classroom and interpret or apply flexibility within the IREAD-3 grade retention policy in significantly different ways. This issue of variability in implementation of policy is widespread and can be observed in many settings and on many topics with notable effect (Cohen, Loeb, Miller, & Wyckoff, 2020; Fowler, 2013; Hall & Chapman, 2018). Specifically, in the realm of decisions regarding grade retention, there are times when administrators and teachers are faced with an ethical dilemma of not following policy in the face of mitigating circumstances or underlying factors of failure for individual students (Larsen & Akmal, 2007). In the dataset used for this study, there was no way of identifying specifically which students were retained according to the IREAD-3 policy or what other determinations

were made locally according to the policy. The policy in Indiana provides a fairly large amount of leeway for districts and schools to make local decisions. Using the available data, I was only able to determine that there was a retention due to the student being reported in the same grade the following year.

Many potential factors, such as the specifics of the IREAD-3 assessment and the cut score used in the retention policy, preexisting interventions and other supports available in schools around the state, and the ISTEP+ test itself as a measure of student achievement, are all particular to Indiana. Consequently, there is some limitation to the ability to generalize the findings of this study to grade retention on the whole. However, as educational standards and policies are specific to particular regions, this is an issue with almost all studies of this kind and narrows the appropriate research questions that can be answered (Wu, West, & Hughes, 2008).

Summary

In this chapter, I reviewed the design of this research study on the academic effects of test-based grade retention. A description of the dataset including the demographics of the sample dataset was presented. I discussed the research questions and the independent and dependent variables used in the statistical analysis. An overview of analytical methods used was provided as well. In the next chapter, I present my statistical analysis and findings.

CHAPTER 4

RESULTS

In this chapter, I begin with a restatement of the purpose of this study and the research questions being addressed. I then describe the dataset that I analyzed and some of the specific student-level factors included. Next, I share descriptive results regarding student performance on ISTEP+ in subsequent years for the 2012 Grade 3 Cohort in Indiana. Then I present inferential statistical analysis on the differences between the performance of the students who did not pass IREAD-3 and were either retained or promoted. This includes t-tests and analysis using longitudinal linear mixed modeling. The chapter concludes with a summary of the various approaches used in the analysis and a comparison of the outcomes.

Purpose of the Study

The purpose of this study is to analyze the effectiveness of test-based grade retention as a policy and, specifically, to analyze the effectiveness of the IREAD-3 policy in the state of Indiana. The focus of this analysis is on students' academic achievement rather than psychological or social effects. Student testing data from the Indiana Department of Education will be used to examine the effects of grade retention in third grade in accordance with the IREAD-3 policy on the achievement levels of those students on the annually administered Indiana Statewide Testing for Educational Progress Plus (ISTEP+) in the years after the decision to retain the student in grade. Test data from Grades 3-8 will be used in analysis. The independent variables are the IREAD-3 score and the decision to retain or not. The dependent variables are ISTEP+ scores in Grades 4-8. The following variables will be controlled for as they have been shown to be associated with student achievement levels as measured by standardized tests: free/reduced lunch status (a measure of socioeconomic status), IEP

(Individualized Education Plan) / disability exemption status, ELL (English Language Learner) status, ethnicity, and gender.

Research Questions

While this study contributes to an understanding of the effectiveness of test-based retention policies in general, the specific research questions of this study relate to effects on student academic achievement of the IREAD-3 policy in the state of Indiana. This study has one primary research question and two ancillary questions that will directly inform the primary question.

> Primary Research Question: What is the relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?
>
> Subquestion 1: What is the relationship between grade retention and student achievement at each grade level 4-8 after the decision to retain?
>
> Subquestion 2: What is the relationship between grade retention and student achievement considered longitudinally across grade levels 4-8 after the decision to retain?

The choice to use grades 4-8 after the decision to retain arises from the current grade level where retention is being considered (Grade 3) and the data available. This range of grades is also consistent with similar studies. Significantly for the purpose of using statewide data in Indiana, after the decision to retain in Grade 3 there is continuous data from the ISTEP+ assessment through Grade 8 over the years of 2012-2018. This also follows the students through to the beginning of high school.

Description of the Dataset

As presented in the previous chapter, the dataset for this study came from a data release request from the Indiana Department of Education with a focus on the 2012 Grade 3 cohort as they progressed through Grade 8. This includes data over the years of 2012-2018. Descriptions are provided in this chapter about how the dataset was "cleaned" to include only complete values for the purpose of running statistical analyses. For the longitudinal multilevel modeling analysis, the dataset was narrowed down to include only students who Did Not Pass IREAD-3 in 2012 in order to do a direct comparison of the effect of being retained or promoted from among this group.

Specific Student-Level Factors of the Dataset

Student mobility has been observed to have a negative impact on student achievement (Burkam, Lee, & Dwyer, 2009; National Research Council and Institute of Medicine, 2010; Mehana & Reynolds, 2004; Reynolds, Chen, & Herbers, 2009; Welsh, 2017), and students who change schools are disproportionately more likely to be from low-income families, have Individualized Education Plans, and be English Language Learners (Ashby, 2010; Welsh, 2017). The negative effect on student achievement even extends to other students in the school (Rumberger, 2003). These findings make it important to consider student mobility in this kind of student academic performance data analysis.

Additionally, policies and practices of grade retention can vary from school to school and district to district. In this dataset, students who attended more than one school per year were duplicated as cases with the only differences in the cases being the Corporation, School, and Days Attended. In order to avoid duplicate values for students who attended more than one school, which would confuse the results of the statistical analysis, I had to consolidate the cases

for a single student in each year. I summed the total number of days attending school in all schools for a given year and kept the Corporation and School data for the school that had the highest number of days attended in the year. This approach allowed me to control for attendance rate on the whole of the academic year. I included the school corporation/district ID to control for local level factors related to policies and approaches they have regarding remediation and retention. I calculated a Student Mobility variable for any students who changed schools in each year. I used this mobility variable in the later analysis.

As an example of the relative prevalence of student mobility for this dataset, I will provide a description of the first year considered in this study. On one hand, of the 85,511 students in the 2012 Grade 3 Cohort, 79,997 (93.6%) did not change schools in Indiana during that academic year. This figure does not account for students who might have moved from out of state or transferred from a private school that did not report data to the Indiana Department of Education. On the other hand, 4859 (5.7%) students changed schools once, 650 (0.7%) students changed schools twice, 80 (0.09%) students changed schools three times, 19 (0.02%) students changed schools four times, and 6 (0.007%) students changed schools 5 times as reported during the 2012 academic year.

Student mobility played a role in excluding students from the data analysis as well. For example, if a student moved out of state for any year of school within the study, 2012-2018, they would be missing a year of assessment scores and would not be included in the final dataset. Once I cleaned the dataset by removing students who had missing data or who were not reported in all Grades 3-8, the other demographic ratios were affected. This can be seen in Tables 4 and 5. While there was some loss of comprehensiveness in this step, the sample sizes used for all statistical analyses were still large due to the use of statewide data.

The IREAD-3 retention policy allows for schools to give Good Cause Exemptions to students through one of three ways: IEP committee decision, ELL committee decision, or having already been retained twice. There was no way to tell which students received them from the released dataset. There are fields for IEP Exemption and ELL Exemption which designate which students were reported to have received an exemption based on having an IEP or ELL designation. These students made up a small proportion of the entire 2012 Grade 3 Cohort, 5.5% and 1.5% respectively, and varied in significant ways from the population at large. The disproportionate representations for IEP students that stand out the most are the areas of DNP IREAD-3, Retained, and Black. The areas that stand out as most disproportionate for ELL students are DNP IREAD-3, Asian, and Hispanic. The full comparison of the breakdowns of these subgroups can be seen in Table 4.

If the state's IREAD-3 policy were followed precisely, every student in Grade 3 would take IREAD-3, and any students who did not pass IREAD-3 would have to retake it the following year even if they were promoted on to Grade 4. The only exceptions would be if a student received a Good Cause Exemption. However, the data on the 2012 Grade 3 cohort shows that only 93 students retook the IREAD-3 in 2013, and by 2014, only 1 student took the IREAD-3. That student was reported in Grade 3 in 2014 and did earn a passing score. Of the 7355 (8.6%) students in the 2012 Grade 3 cohort who did not pass the IREAD-3, there were 4682 (5.5%) IEP Exemptions and 1240 (1.5%) ELL Exemptions. That would leave 1433 students who would have been retained according to the IREAD-3 policy. Of these 1433 students, 262 students were not reported by any Indiana school in 2013. That would leave 1171 students to be retained. By my calculations of what was reported in the dataset, 2760 (3.2%) students were retained in Grade 3 for 2013. Ostensibly, there were 1589 students retained for

Table 4

IEP and ELL Exemptions Relative to the Entire 2012 Grade 3 Cohort

Student Category	IEP Exemption N (% of category)	IEP Exemption after "cleaning" N (% of category)	ELL Exemption N (% of category)	ELL Exemption after "cleaning" N (% of category)
All Students	4682 (5.5%)	3064 (4.8%)	1240 (1.5%)	929 (1.5%)
DNP IREAD-3	4138 (56.3%)	2765 (61.8%)	1096 (14.9%)	822 (18.4%)
Retained in Grade 3	316 (11.4%)	151 (11.8%)	47 (1.7%)	31 (2.4%)
Female	1671 (4.0%)	1120 (3.6%)	559 (1.3%)	415 (1.3%)
Male	3011 (6.8%)	1944 (6.0%)	681 (1.5%)	514 (1.6%)
American Indian	9 (4.5%)	6 (4.1%)	2 (1.0%)	2 (1.4%)
Asian	36 (2.1%)	22 (2.0%)	184 (10.9%)	120 (10.9%)
Black	1067 (10.6%)	655 (10.2%)	33 (0.3%)	21 (0.3%)
Hispanic	348 (3.9%)	222 (3.4%)	937 (10.6%)	737 (11.2%)
Multiracial	246 (6.2%)	164 (5.8%)	11 (0.3%)	9 (0.3%)
Native Hawaiian or Other Pacific Islander	1 (2.4%)	0 (0%)	1 (2.4%)	1 (4.0%)
White	2975 (4.9%)	1995 (4.3%)	72 (0.1%)	39 (0.1%)
Free / Reduced meals	3457 (8.4%)	2284 (7.3%)	1005 (2.4%)	847 (2.7%)
Paid meals	1091 (2.9%)	780 (2.4%)	135 (0.4%)	82 (0.3%)
Unknown pay status	134 (2.1%)	0 (0%)	100 (1.6%)	0 (0%)

Notes: The percentages are of the total for that group in the 2012 Grade 3 Cohort. For example, the 4138 students with IEP who Did Not Pass the IREAD-3 made up 56.3% of the students who Did Not Pass IREAD-3.

Table 5

Reported Grade in 2013 for the 2012 Grade 3 Cohort

	N	%
Grade 3	2760	3.2%
Grade 4	79794	93.3%
Grade 5	76	0.1%
Grade 6	7	0.0%
Grade 7	2	0.0%
Not Reported	2872	3.4%

Notes. The Not Reported category includes those students not reported by any schools in 2013, presumably due to transfer to private schools or out of state but could also include drop-outs or deaths.

other reasons than following the IREAD-3 policy. The reported grade levels for the 2013 academic year are listed in Table 5. There were 5285 (6.1%) students who did not pass IREAD-3 and were promoted to Grade 4 for 2013. There were 2070 (2.4%) students who did not pass IREAD-3 and were retained in Grade 3 for 2013. This shows that some students who could have been promoted due to IEP or ELL statuses were still retained rather than be promoted through the use of Good Cause Exemptions provided for as a part of the IREAD-3 policy in Indiana.

Retention and Promotion in the Dataset

Once I had merged the years together into a single file, I calculated some additional variables to assist with my analysis. I calculated a binary variable for DNPIREAD3 which was any student who scored less than 446 on the IREAD-3 test in 2012. I also added a binary variable for Retained for any student who was reported in Grade 3 again in 2013. Using these variables and some other conditions, I calculated grade level specific test scores for math and English/Language Arts accounting for whether the student was retained to allow for same-grade

comparisons of student achievement using If statements. This was done for all four of the groups of the 2012 Grade 3 cohort as follows: DNP and Retained, DNP and Promoted, Pass and Retained, and Pass and Promoted. This comparison is presented after a brief discussion about the demographics of students who DNP IREAD-3 and/or were retained.

The overall demographics for the 2012 Grade 3 Cohort are presented in the previous chapter. Since the focus of this study revolves around students who did not pass IREAD-3 who were then retained or promoted, I begin by examining the student demographic breakdown of these subgroups of students. The data in Table 6 are organized into three subgroups: Did Not Pass IREAD-3, Retained in 3rd Grade, and the combination of the two for students who Did Not Pass IREAD-3 and were retained

Before running further data analysis, I cleaned the dataset by removing all students with missing or unknown values in Free/Reduced status, Math ISTEP+ score, or ELA ISTEP+ score or who did not have data reported in all school years from 2012 to 2017, or 2018 in the case of the retained students. This ensured that the remaining students had ISTEP+ scores for all grades from Grade 3 to Grade 8 and that they had only been retained at most once in that span—only after Grade 3. The distribution of many of these demographic factors was altered, as can be seen in Table 7.

Since the ISTEP+ assessment changed in 2015, I could not appropriately correlate the scale scores from 2012-2018. By standardizing the test scores for each year, I hoped to account for this potential discrepancy. This approach of comparing individuals to the entire population was especially appropriate since part of the underlying reasoning for grade retention as a practice is to allow students to “catch up” to their peers by giving them more time to master concepts and skills. I will note that in order to calculate the z-score for each academic year, scores from all

Table 6

Demographics Breakdown of Did Not Pass on IREAD-3 and/or Retained in 3rd Grade for 2012 Cohort

Student Category	Did Not Pass IREAD-3 N (% of category)	Retained in 3rd Grade N (% of category)	Did Not Pass IREAD-3 and Retained N (% of category)
All Students	7355 (8.6%)	2760 (3.2%)	2070 (2.4%)
Female	2855 (6.9%)	1208 (2.9%)	894 (2.2%)
Male	4500 (10.2%)	1552 (3.5%)	1176 (2.7%)
American Indian	11 (5.4%)	2 (1.0%)	2 (1.0%)
Asian	214 (12.7%)	14 (0.8%)	9 (0.5%)
Black	1922 (19.0%)	1007 (10.0%)	847 (8.4%)
Hispanic	1340 (15.1%)	301 (3.4%)	229 (2.6%)
Multiracial	375 (9.4%)	165 (4.1%)	127 (3.2%)
Native Hawaiian or Other Pacific Islander	2 (4.8%)	0 (0%)	0 (0%)
White	3491 (5.8%)	1271 (2.1%)	856 (1.4%)
Free / Reduced meals	5628 (13.6%)	2103 (5.1%)	1653 (4.0%)
Paid meals	1367 (3.6%)	468 (1.2%)	300 (0.8%)
Unknown pay status	360 (5.6%)	189 (3.0%)	117 (1.8%)

grades from 3rd to 8th were included together. While the ISTEP+ scale score cut values increase roughly 20-30 points along with each progression from grade level to grade level (IDOE, n.d.), there is not a tight correlation that would allow for direct comparison of scores for students from one grade to the next as might exist in other year-to-year achievement tests, such as the widely used NWEA MAP Growth assessments. Since my purpose was to compare performance in

Table 7

Demographics Breakdown of Did Not Pass on IREAD-3 and/or Retained in 3rd Grade for 2012 Cohort

Student Category	Did Not Pass IREAD-3 N (% of category)	Retained in 3rd Grade N (% of category)	Did Not Pass IREAD-3 and Retained N (% of category)
All Students	4473 (7.1%)	1275 (2.0%)	949 (1.5%)
Female	1745 (5.6%)	561 (1.8%)	416 (1.3%)
Male	2728 (8.5%)	714 (2.2%)	533 (1.7%)
American Indian	8 (5.4%)	1 (0.7%)	1 (0.7%)
Asian	139 (12.6%)	6 (0.5%)	4 (0.4%)
Black	944 (14.6%)	382 (5.9%)	313 (4.9%)
Hispanic	953 (14.5%)	162 (2.5%)	129 (2.0%)
Multiracial	219 (7.7%)	77 (2.7%)	58 (2.1%)
Native Hawaiian or Other Pacific Islander	1 (4.0%)	0 (0%)	0 (0%)
White	2209 (4.8%)	647 (1.4%)	444 (1.0%)
Free / Reduced meals	3580 (11.4%)	1032 (3.3%)	798 (2.5%)
Paid meals	893 (2.8%)	893 (2.8%)	151 (0.5%)
Unknown pay status	0 (0%)	0 (0%)	0 (0%)

relation to peers each year, the difference from this calculation and a z-score for just the individual grade level should not impact such an analysis, but it does mean that there would be lower expected z-score values in lower grades. For example, for the 74,791 students who earned Pass on IREAD-3 in 2012 and were promoted to Grade 4 for 2013, the mean of the z-scores on their Grade 3 English / Language Arts test was 0.119, the median was 0.064, and the

standard deviation was 0.885. In comparison, the 67,165 students who remained out of that same 2012 Grade 3 cohort and took the ISTEP+ Grade 8 English / Language Arts test had mean z-scores of 0.837 with a median of 0.850 and standard deviation of 0.929. This across grades standardizing seemed to be the best option for working with the year-to-year ISTEP+ scores in this study, especially since the test changed in 2015 and the study is being performed on a same-grade comparison. So, the z-scores should be expected to increase from grade to grade, but there is no direct correlation table. They are used on a peer comparison basis.

Population Subgroup Comparisons

As an initial analysis of the effect that retention might have on academic achievement and following in the approach of Lorence (2014), I calculated averages for specific subgroups of interest concerning the IREAD-3 test and retention. I created four groups of students, as seen in Table 9, based on two specific factors of concern: receiving a Pass or Did Not Pass on IREAD-3 and being retained or promoted. I only included students who were retained in Grade 3 once and who were reported in the correct subsequent grade level for a given academic year. This last stipulation was to try to account for the discrepancies in reported grade level that could be seen in each year of the reported data. It also filters out students who might have been retained an additional time. This results in decreasing numbers of students in each subgroup as the grade-levels increase.

The results in this table show that students who did not pass IREAD-3 and were still promoted performed much more poorly on subsequent years of ISTEP+ ELA than students who were retained. For example, in Grade 8, the mean z-score for students of the DNP/Promoted group was -.802 while the mean z-score of the DNP/Retained was -.287. However, it is important to note that the first group of students who did not pass IREAD-3 and were promoted

includes many of the students who received Good Cause Exemptions. This could mean that students who might have the most significant challenges to learning and achieving are included here. On the other hand, many of those students could have taken alternative assessments that can be determined to be a better fit for students with an IEP. In Indiana, the IMAST was given through 2014 for certain students in grades 3-8 who were expected to graduate with a regular high school diploma. The proportions of all students who took the IMAST rather than the ISTEP+ from 2012 to 2014 were 4.2%, 3.9% and 3.8%, respectively. While the assessment was significantly modified from the ISTEP+, the scale scores were intended to be aligned. These scores were reported in the data files I received for 2012-2014. If anything, students who took the IMAST and had those reported for their math and ELA test scores would have scored higher than if they had taken the ISTEP+, so there could be a slight inflation of the scores of the group that included DNP IREAD-3 and Promoted. The ISTAR assessment was intended for students not on diploma track for the purpose of documenting progress and was replaced by the I AM assessment in Indiana for the Spring 2019 testing window (IDOE, n.d.). Presumably, some of the unreported ISTEP+ scores in the datasets would be for students who took the ISTAR. This should not affect the averages for any of the groups as they would not be calculated with missing data.

The data in Tables 8 and 9 are for all valid values. While this information carries some interest at a policy level as it is the most comprehensive view, it is important to remember the muddiness to the water that is brought by not accounting for so many factors such as students with more severe disabilities, non-English speaking backgrounds, etc. Still, it is notable how, among the students who DNP IREAD-3, retained students consistently performed better than their promoted peers. Also, the growth of retained students, whether they passed IREAD-3 or

not, was greater from Grades 3-8 than even the students who passed IREAD-3 and were promoted while more than doubling the growth of the students who did not pass IREAD-3 and were promoted.

While I sought to keep as many students included as possible in the early stages of analysis, later statistical computations required a "cleaner" dataset. For comparison sake, I went ahead and performed the same subgroup comparison on that further cleaned dataset. I removed all students with missing or unknown values in Free/Reduced status, Math ISTEP+ score, or ELA ISTEP+ score or who did not have data reported in all school years from 2012 to 2017, or 2018 in the case of the retained students. The data in Tables 17 and 18 in Appendix D represent the same subgroup comparisons found in Tables 8 and 9 but are made on that smaller, cleaned dataset. My cleaned dataset consisted of 59454 students for whom there was complete data, including both ELA and Math scores, for grades 3-8 and who were either never retained or only retained in 3rd grade. If a student was retained a second or third time from grades 3-8, they were removed. Of this sample, 55802 (93.9%) passed IREAD-3 in 2012. Additional demographic information on this group includes: 29874 (50.2%) were male and 29580 (49.8%) were female; 140 (0.2%) were American Indian, 1048 (1.8%) were Asian, 5813 (9.8%) were Black, 6182 (10.4%) were Hispanic, 2641 (4.4%) were Multiracial, 25 (0.0%) were Native Hawaiian or Other Pacific Islander, and 43605 (73.3%) were White; 29011 (48.8%) were Free/Reduced and 30443 (51.2%) were Paid.

Table 8

Same-Grade Comparison of ISTEP+ English / Language Arts Standardized Test Results (as z-scores) for 2012 Grade 3 Cohort with All Valid Values Included

Grade	Academic Year		Did Not Pass IREAD-3 and Promoted	Did Not Pass IREAD-3 and Retained	Passed IREAD-3 and Retained	Passed IREAD-3 and Promoted
Grade 3	2012	N	4935	1986	554	74791
		Mean	-1.148	-1.641	-1.110	.119
		Median	-1.299	-1.585	-1.078	.064
		S. D.	1.327	.885	.704	.885
Grade 3 (post-retention)	2013	N	---	1936	634	---
		Mean	---	-1.285	-.986	---
		Median	---	-1.315	-1.046	---
		S. D.	---	.691	.593	---
Grade 4	2013 or 2014	N	4614	1323	549	72548
		Mean	-1.190	-1.016	-.737	-.250
		Median	-1.173	-1.032	-.766	-.329
		S. D.	1.054	.666	.629	.870
Grade 5	2014 or 2015	N	4420	1353	549	71109
		Mean	-.858	-.561	-.296	.005
		Median	-.860	-.436	-.227	-.017
		S. D.	.859	.820	.715	.697
Grade 6	2015 or 2016	N	4391	1276	522	70235
		Mean	-.775	-.631	-.240	.267
		Median	-.666	-.631	-.217	.332
		S. D.	.853	.620	.676	.759
Grade 7	2016 or 2017	N	4063	1203	500	68233
		Mean	-.880	-.500	-.122	.609
		Median	-.853	-.480	-.140	.612
		S. D.	.731	.615	.707	.814
Grade 8	2017 or 2018	N	3942	1172	500	67165
		Mean	-.802	-.287	-.024	.837
		Median	-.806	-.139	.136	.850
		S. D.	.835	1.073	1.130	.929

Table 9

Same-Grade Comparison of ISTEP+ Mathematics Standardized Test Results (as z-scores) for 2012 Grade 3 Cohort with All Valid Values Included

Grade	Academic Year		Did Not Pass IREAD-3 and Promoted	Did Not Pass IREAD-3 and Retained	Passed IREAD-3 and Retained	Passed IREAD-3 and Promoted
Grade 3	2012	N	5106	2031	565	74447
		Mean	-.887	-1.410	-1.145	.108
		Median	-.945	-1.415	-1.173	.102
		S. D.	1.174	.863	.786	.924
Grade 3 (post-retention)	2013	N	---	1942	639	---
		Mean	---	-1.292	-1.076	---
		Median	---	-1.343	-1.089	---
		S. D.	---	.810	.721	---
Grade 4	2013 or 2014	N	4665	1327	554	72475
		Mean	-.999	-1.016	-.766	-.269
		Median	-.936	-1.032	-.803	-.326
		S. D.	.871	.666	.655	.890
Grade 5	2014 or 2015	N	4473	1353	549	71123
		Mean	-.754	-.418	-.304	.106
		Median	-.676	-.306	-.241	.099
		S. D.	.828	.770	.726	.766
Grade 6	2015 or 2016	N	4391	1271	526	70170
		Mean	-.573	-.367	-.224	.312
		Median	-.470	-.349	-.217	.371
		S. D.	.877	.600	.595	.731
Grade 7	2016 or 2017	N	4089	1209	498	67812
		Mean	-.286	-.157	-.019	.601
		Median	-.289	-.149	-.018	.579
		S. D.	.570	.555	.571	.693
Grade 8	2017 or 2018	N	3980	1173	500	66932
		Mean	-.107	.019	.179	.929
		Median	-.076	.213	.378	.946
		S. D.	.724	1.089	1.078	.690

Also, 2905 (4.9%) students changed schools at least once during the 2012 academic year. The proportion of students who moved between schools during each year was 4.4%, 3.8%, 3.8%, 3.7%, 3.8%, and 8.7%. The only students included in the 2018 academic year were students retained in Grade 3 in 2012. There were 1112 (1.9%) students who were retained, and of those, 97 (8.7%) changed schools during the 2018 school year.

T-test Comparisons of Retained and Promoted

The size of the dataset and its quality of including all students in Indiana make the previous tables applicable for policy considerations from a descriptive statistical standpoint. However, even with large samples, it is important to consider whether observed differences are actually statistically significant. It is also helpful to consider effect sizes to better understand the magnitude of the effect of a factor. Therefore, after removing any students with missing data for any of the variables, I conducted a t-test to compare the means of the ISTEP+ scores for DNP and Promoted against DNP and Retained to ascertain whether these groups of students actually performed differently. There were 3652 students in the 2012 Grade 3 Cohort who Did Not Pass IREAD-3. Table 10 shows the mean tests scores for each group, the results of an independent samples t-test, and the effect size of the difference between the means as Cohen's d. This comparison was made same-grade, so the retained group has ISTEP+ scores from a year later than the promoted group. The exception is for Grade 3 where the first attempt of the retained group was used rather than the scores in the year following retention since that would seem to put them at a significant advantage. As can be seen in Table 13, the students who were retained after they did not pass IREAD-3 in 2012 performed worse in the initial Grade 3 assessment prior to retention but better in every other grade level than their peers who were promoted. This was true for both English/Language Arts and mathematics ISTEP+ performance.

The differences at each grade level between the two groups, promoted and retained, were all statistically significant ($p<.001$). The negative effect size for Grade 3 shows that the retained students scored lower in that year. However, the retained students scored higher than the promoted students in every grade post-retention. Following the guidelines laid out by Cohen (1988), the effect sizes for ELA range from small in Grade 4 (d = .253) and Grade 6 (d =.044) to medium in Grade 5 (d =.672) and Grade 7 (d =.578) to large in Grade 8 (d =1.044). The trend of effects were similar for Math with effect sizes ranging from small in Grade 4 (d =.146), Grade 6 (d =.208), and Grade 7 (d =.313) to medium in Grade 5 (d =.624) and Grade 8 (d =.536).

Since the IREAD-3 policy specifically allows for the Good Cause Exemptions for students with IEPs or who are English Language Learners, I then removed all students who had a reported IEP exemption or ELL exemption. The results of the t-tests for these 758 students are listed in Table 11. Overall, there is a clear trend of statistically significant differences with the retained students performing better than their promoted peers. It is notable that for this group of students, the retained students did not perform lower than their promoted peers in Grade 3, prior to the decision to retain. The differences in ELA scores between the two groups are significant for every grade except 4th with effect sizes ranging from small for Grade 6 (d =.044) to medium for Grade 3 (d =.442) and Grade 7 (d =.578) to large for Grade 5 (d = 1.108) and Grade 8 (d = 1.045). The differences in the Math scores between the two groups are only significant in Grades 5, 6, and 8 with effect sizes ranging from small for Grade 6 (d =.146) to large for Grade 5 (d =.909) and Grade 8 (d =.782). While the results vary somewhat from the data that included the IEP and ELL exemptions, there is still a clear trend that the retained students outperformed their promoted peers annually and all the way through Grade 8.

Table 10

Same-Grade t-test Comparison of Students who Did Not Pass IREAD-3 in 2012 (Only Students With Complete Data)

Grade	ISTEP+ Subject		N	Mean	SD	t-test	Cohen's d
Grade 3	ELA	Promoted	2845	-1.157	1.257	t(2654) = -2.78,	-.079*
		Retained	807	-1.248	.637	p=.049	
	Math	Promoted	2845	-.864	1.121	t(1871) = -10.15,	-.331***
		Retained	807	-1.213	.771	p<.001	
Grade 4	ELA	Promoted	2845	-1.188	1.004	t(2056) = 7.27,	.253***
		Retained	807	-.975	.636	p<.001	
	Math	Promoted	2845	-.991	.851	t(1604) = 3.91,	.146***
		Retained	807	-.879	.676	p<.001	
Grade 5	ELA	Promoted	2845	-.846	.820	t(3040) = 21.45,	.672***
		Retained	807	-.420	.361	p<.001	
	Math	Promoted	2845	-.719	.803	t(2348) = 18.56,	.624***
		Retained	807	-.312	.454	p<.001	
Grade 6	ELA	Promoted	2845	-.627	.430	t(1049) = 1.03,	.044***
		Retained	807	-.604	.606	p<.001	
	Math	Promoted	2845	-.460	.523	t(1165) = 5.00,	.208***
		Retained	807	-.342	.608	p<.001	
Grade 7	ELA	Promoted	2845	-.859	.708	t(1496) = 15.18,	.578***
		Retained	807	-.479	.603	p<.001	
	Math	Promoted	2845	-.280	.547	t(1327) = 7.88,	.313***
		Retained	807	-.111	.533	p<.001	
Grade 8	ELA	Promoted	2845	-.785	.815	t(2205) = 30.52,	1.044***
		Retained	807	-.084	.486	p<.001	
	Math	Promoted	2845	-.108	.708	t(1848) = 14.65, p<.001	.536***
		Retained	807	.219	.492		

Notes. *p<.05; **p<.01; ***p<.001; effect sizes are positive when Retained is greater than Promoted

Table 11

Same-Grade t-test Comparison of Students who Did Not Pass IREAD-3 in 2012 (Only Students With Complete Data and no IEP or ELL exemptions)

Grade	ISTEP+ Subject		N	Mean	SD	t-test	Cohen's d
Grade 3	ELA	Promoted	92	-1.477	.847	t(100) = 2.75,	.442**
		Retained	666	-1.227	.513	p=.007	
	Math	Promoted	92	-1.287	.797	t(756) = 0.80,	.089
		Retained	666	-1.223	.716	p=.423	
Grade 4	ELA	Promoted	92	-1.196	.630	t(756) = 3.51,	.371
		Retained	666	-.976	.555	p=.653	
	Math	Promoted	92	-1.257	.704	t(756) = 5.10,	.547
		Retained	666	-.888	.643	p=.196	
Grade 5	ELA	Promoted	92	-.935	.655	t(97) = 8.25,	1.108***
		Retained	666	-.362	.326	p<.001	
	Math	Promoted	92	-.844	.761	t(100) = 6.98,	.909***
		Retained	666	-.278	.443	p<.001	
Grade 6	ELA	Promoted	92	-.627	.430	t(1049) = 1.03,	.044***
		Retained	666	-.604	.606	p<.001	
	Math	Promoted	92	-.369	.416	t(149) = 1.53,	.146***
		Retained	666	-.294	.598	p<.001	
Grade 7	ELA	Promoted	92	-.859	.708	t(1496) = 15.18,	.578***
		Retained	666	-.479	.603	p<.001	
	Math	Promoted	92	-.388	.466	t(756) = 5.31,	.619
		Retained	666	-.079	.530	p=.369	
Grade 8	ELA	Promoted	92	-.785	.815	t(2205) = 30.52,	1.045***
		Retained	666	-.084	.486	p<.001	
	Math	Promoted	92	-.166	.592	t(107) = 6.50,	.782**
		Retained	666	.252	.470	p=.002	

Notes. *p<.05; **p<.01; ***p<.001; effect sizes are positive when Retained is greater than Promoted

Table 12

Demographic Breakdown for Treatment Dataset (Students who DNP IREAD-3)

Demographic Category	% of Dataset
Female	41.1%
Male	58.9%
Another Ethnicity	3.1%
Black	20.5%
Hispanic	22.5%
Multiracial	4.8%
White	49.0%
Free / Reduced meals	77.2%
Paid meals	22.8%
Changed Schools	7.8%
Remained in Same School All Year	92.2%
ELL	19.3%
IEP	58.5%
Retained	23.1%
Promoted	76.9%

While these t-tests specifically target the factor of retention, they do not control for any additional factors but rather just provide a direct comparison of values between two groups. So,

while the t-tests showed that the differences in the mean scores of the students who were retained or promoted were statistically significant, I wanted to do an additional analysis of individual students over time that would control for factors that have been linked to student achievement in other studies and that I had access to in this dataset. The final portion of analysis performed through the use of longitudinal multilevel modeling is the most important to consider.

Longitudinal Analysis

As a final step of analysis, I organized the dataset to only include students who Did Not Pass IREAD-3 and who had no missing data for grades 3-8. This made up the "Treatment" dataset. The demographic make-up of this dataset can be seen in Table 12. In this final group, 23.1% of the students who did not pass IREAD-3 were actually retained, and this retention was considered as the treatment variable. Since the focus of this study is on the academic effects of grade retention for students who did not pass IREAD-3, I primarily consider the difference between the retained and promoted students in these models while all other factors were included for the purpose of controlling for their effects. I calculated Time as shifted values of Grade so that Grade 3 was equal to Time 0, etc. After running some sequential tests to see if time, time squared, and/or time cubed were significant factors on their own, I was able to determine that both time and time squared were significant factors while time cubed was not. However, due to multicollinearity concerns between Time and $(\text{Time})^2$, I ended up only using Time as a factor. One model was run for the ELA z-scores and one was run for the Mathematics z-scores. The model utilized for this study used the following variables:

$$Y_{ig} = \beta_{0g} + \beta_{Retained(i)} + \beta_{Gender(i)} + \beta_{Black(i)} + \beta_{Hispanic(i)} + \beta_{Multiracial(i)} + \beta_{AnotherEthnicity(i)} + \beta_{Move(ig)} + \beta_{SES(g)} + \beta_{IEP(i)} + \beta_{ELL(i)} + \beta_{Age2012(i)} + \beta_{AttendedDays(g)} + \beta_{Ziread2012(i)} + \beta_{Time(g)} + \beta_{CorpID/StudentID(g)} + e_{ig}$$

The values of each of the variables are described in the previous chapter. The *i* and *g* stand for initial values in 2012 (*i*) and values that vary as the student moves from grade level to grade level (*g*). The CorpID/StudentID term is the nested factor of the student within the school corporation. The e_{ig} term represents error in estimating the slope of the model. The model was run using the lmer function the lme4 library of R.

Different school districts can vary greatly in their policies and practices regarding all retention in general and the IREAD-3 policy in particular. Besides differences in the actual decisions of whether or not to retain, there can be a great deal of variance in the availability and extensiveness of additional supports and interventions that may accompany these decisions. Due to these concerns over the amount of variation that can exist between school districts, the final analysis I used was longitudinal multilevel modeling nesting students within their school corporations (which is what school districts in Indiana are called). I was able to use the lmer function in the R statistical library lme4 to produce such a model.

In the ELA model, seven factors were statistically significant: Retained, Time, Gender, Age in Grade 3, Move, Days Attended, and IREAD-3 z-score. Seeing this was the case and being surprised that some additional factors such as SES were not significant, I investigated the possibility of collinearity among the numerous factors included in the models. I completed multicollinearity tests, and no Variance Inflation Factor (VIF) was large enough to indicate collinearity, except the Time and $(\text{Time})^2$ variables. There would seem to be a relationship between SES and the other variables that is causing the model to produce this result, and I am unable to explain that SES would not be significant factor of student achievement along with some of the other demographic elements used in the model, as it often is often found to be in analyses of student achievement (Grissmer, Ober, & Beekman, 2014; Skiba & Rausch, 2004).

However, this could be explained by the fact that this dataset only has a very simplistic measure for SES based on Free/Reduced Lunch status. The results for ISTEP+ ELA z-scores in the final longitudinal multilevel model used are provided in Table 13.

While the focus of this study is on the effect of retention on subsequent academic performance, which is discussed further in the next chapter, a brief description of other

Table 13

Estimate of Fixed Effects for Multilevel Modeling on ISTEP+ English / Language Arts Z-scores

Parameter	Estimate	Std. Error	t-value	95% Conf. Interval	
				Lower Bound	Upper Bound
(Intercept)	-0.561*	0.170	-3.297	-0.895	-0.228
Retained	0.113*	0.030	3.769	0.054	0.172
Time	0.126*	0.003	43.823	0.121	0.132
Age in 3rd Grade	-0.041*	0.014	-3.007	-0.069	-0.014
Gender	0.139*	0.014	9.658	0.111	0.167
Black	-0.037	0.022	-1.704	-0.080	0.006
Hispanic	-0.002	0.028	-0.073	-0.056	0.052
Multiracial	0.012	0.034	0.352	-0.054	0.078
Another Ethnicity	0.067	0.049	1.369	-0.029	0.163
Move	-0.077*	0.020	-3.880	-0.115	-0.038
Days Attended	0.001*	0.001	1.825	0.000	0.002
SES	-0.015	0.015	-0.974	-0.045	0.015
IEP	-0.027	0.031	-0.871	-0.087	0.034
ELL	-0.073	0.040	-1.821	-0.151	0.005
Z-score of IREAD-3	0.240*	0.012	20.210	0.217	0.264

*Notes. * The estimated effect is significant at the .05 level.*

significant factors in the model is provided here. Time as a positive factor is likely understood as a function of the increasing scale scores on ISTEP+ and how the year-to-year z-scores were calculated across grades 3-8 as I described earlier. That Age in Grade 3 is negative and significant means scores were lower the older a student was in Grade 3 for the first time. This likely correlates with students who may have been previously retained in grades K-3 prior to 2012 and was predictable; however, I did not have any data on those prior school years. Gender was coded with female being 1, so a positive effect (0.139) shows that females scored better than their male peers. Move was dummy-coded as 1 for when a student changed schools within a given academic year, so the negative effect (-0.077) shows that students who did change schools scored lower than their peers who remained in the same school for the entire year. There was no differentiation within this model for students who moved multiple times within a school year. The effect of Days Attended was positive and small (0.001) with estimated effect being tied to single days which indicates that more days in attendance at school correlated to higher scores. Finally, a positive effect (0.240) for the IREAD-3 z-score indicates that better scores on that assessment were related to better scores on ISTEP+ even among this group of students who all scored as Did Not Pass.

The same factors that were significant in the ELA model were also significant in the Math model: Retained (0.094), Time (0.205), Gender (-0.070), Age in Grade 3 (-0.049), Move, Days Attended (0.004), and IREAD-3 z-score (0.177). Additionally, Black (-0.163) and Multiracial (-0.072) were also significant. Again, after running collinearity tests, the only VIF values high enough to indicate multicollinearity were Time and $(Time)^2$. Consequently, it appears that there may be some relationship between SES and other factors, such as IEP and ELL, that are causing one or more of them to not come out as statistically significant in this

model. Due to the collinearity issues, $(Time)^2$ was not used as a factor. The results for the longitudinal multilevel model for Math are presented in Table 14. I will briefly discuss the additional significant factors in this model. Black and Multiracial were the two ethnicities with significant effects. These were both negative--indicating that students of these ethnicities scored lower than their peers.

Table 14

Estimate of Fixed Effects for Multilevel Modeling on ISTEP+ Mathematics Z-scores

Parameter	Estimate	Std. Error	t-value	95% Conf. Interval	
				Lower Bound	Upper Bound
(Intercept)	-0.993*	0.164	-6.058	-1.314	-0.672
Retained	0.094*	0.030	3.092	0.035	0.154
Time	0.205*	0.002	85.395	0.201	0.210
Age in 3rd Grade	-0.049*	0.014	-3.500	-0.076	-0.022
Gender	-0.070*	0.015	-4.783	-0.099	-0.041
Black	-0.163*	0.022	-7.480	-0.206	-0.120
Hispanic	-0.028	0.028	-1.004	-0.084	0.027
Multiracial	-0.072*	0.034	-2.112	-0.140	-0.005
Another Ethnicity	0.077	0.050	1.537	-0.021	0.176
Move	-0.077*	0.017	-4.546	-0.110	-0.044
Days Attended	0.004*	0.001	8.718	0.003	0.005
SES	-0.016	0.014	-1.166	-0.044	0.011
IEP	0.054	0.031	1.710	-0.008	0.115
ELL	-0.003	0.041	-0.085	-0.083	0.076
Z-score of IREAD-3	0.177*	0.012	14.530	0.153	0.200

*Notes. * The estimated effect is significant at the .05 level.*

It would likely be less probable for these students to have been classified as English Language Learners than Hispanic or those included in the Another Ethnicity category (such as Asian), so that could be part of the explanation of why they show as significant while other ethnicities do not.

In summary of the longitudinal multilevel modeling analysis, the effect of being retained was significant and positive for both ELA and Math. The specific comparisons of retained and promoted students for each model are provided in Table 15. These results show that students who did not pass IREAD-3 and were subsequently retained after their first 3rd grade year in 2012 scored on average .113 standard deviations higher on the ISTEP+ ELA assessments than their promoted peers, and they scored on average .094 standard deviations higher on the ISTEP+ Math assessments.

Table 15

Summary of the Effects of Retention in ELA and Math Using Longitudinal Multilevel Modeling with School District Nesting

	Estimate	Std. Error	t-value	95% Conf. Interval	
				Lower Bound	Upper Bound
ELA Z-scores					
Retained	0.113*	0.030	3.769	0.054	0.172
Math Z-scores					
Retained	0.094*	0.030	3.092	0.035	0.154

*Notes. * The mean difference is significant at the .05 level.*

In consideration of the relationships between some of the factors and of the potential for strange reporting regarding the Good Cause Exemptions provided for in the IREAD-3 policy, I performed one additional run of the model. In this version, I removed all students who had IEP

or ELL exemptions from the cleaned dataset and subsequently removed IEP and ELL as factors in the longitudinal multilevel model. Table 19 in Appendix D shows the effects of retention in that model. Notably, the effects of retention on both ELA scores (.212) and Math scores (.180) were even greater in this model.

Summary of Results

Using a same-grade approach for comparison, students who were retained in Grade 3 after failing to pass to IREAD-3 assessment scored better than their promoted peers in subsequent achievement test scores in Grades 4-8 as measured by the ISTEP+ assessment. This held true across grades and subject areas. Additionally, the differences were found to be statistically significant by using t-tests to examine the differences at each grade level. The first part of this analysis considered retention as the only distinguishing factor and did not control for factors that have often been associated with student performance on achievement tests in previous studies. This led to the final part of the statistical analysis using longitudinal multilevel modeling. Of the students who did not pass IREAD-3, the retained students scored better on average than their promoted peers while controlling for other factors such as gender, ethnicity, age, socio-economic status, Individualized Education Plan status, English Language Learner status, mobility between schools, and school attendance. This difference was statistically significant for both English/Language Arts and Mathematics.

Conclusion

This chapter began with a description of the variables in the study and a presentation of descriptive statistics on the dataset including extensive demographic information. Next, an initial analysis of the averages of different subgroups based on IREAD-3 Pass Status and Retention Status was provided. After this categorical comparison, the results of t-tests for

statistical significance of the differences in the means between the subgroups were discussed. Finally, a longitudinal linear mixed model analysis was explained, and the results were presented. In Chapter 5, I provide a summary of this study and discuss its relationship to previous studies on grade retention before considering implications for future research and practice.

CHAPTER 5

CONCLUSIONS

In this chapter, I begin by summarizing this study, including the problem, purpose, and basic design as it relates to the research questions. Next, I provide a discussion of the findings of the study followed by a discussion of implications for practice and policy including recommendations for school administrators as well as policymakers. I then consider where this study fits within the body of prior research and make recommendations for additional research in the future. The chapter ends with concluding remarks about grade retention and student achievement.

Summary of the Study

The purpose of this study was to analyze the effectiveness of test-based grade retention as a policy. The primary focus of this analysis is on academic achievement rather than financial, social, or psychological effects. I used student data released from the State of Indiana to analyze the effectiveness of the IREAD-3 policy in the state of Indiana. The students' academic achievement as measured by ISTEP+ ELA and Math assessments for the 2012 Grade 3 cohort were the outcome variables used while whether the student passed IREAD-3 and was retained in Grade 3 served as the student-level factor of interest. The data analysis includes three parts: a direct comparison of all students by subgroups, t-tests on the differences of the means between subgroups at each successive grade-level through Grade 8, and a longitudinal multilevel modeling analysis across Grades 4-8 controlling for numerous other student-level factors.

Discussion of Findings

In each part of the analysis, there was a positive correlation between retention and subsequent performance on assessments of student achievement when compared to promoted

peers. As might be expected, the positive effect of retention in Grade 3 appeared to become smaller over time but remained positive and significant through Grade 8. Applying the results of the statistical analysis, the answers to the research questions for this study were as follows:

Primary Research Question: What is the relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?

Finding: In a same-grade comparison of students who Did Not Pass IREAD-3, the relationship of retention in Grade 3 to subsequent student achievement performance in English / Language Arts and Mathematics as measured on the ISTEP+ assessment is both significant and positive.

Subquestion 1: What is the relationship between grade retention and student achievement at each grade level 4-8 after the decision to retain?

Finding: In a same-grade comparison of students who Did Not Pass IREAD-3, the students who were retained in Grade 3 performed better on average than their promoted peers. This difference was consistent for both ELA and Mathematics and was almost always statistically significant (exceptions being for non-exemption students in Grade 4 ELA, Grade 4 Math and Grade 7 Math).

Subquestion 2: What is the relationship between grade retention and student achievement considered longitudinally across grade levels 4-8 after the decision to retain?

Finding: In a same-grade comparison of students who Did Not Pass IREAD-3, the students who were retained in Grade 3 performed better on average than their promoted peers through Grade 8. The effect of retention was statistically significant and positive for both ELA (0.113) and Mathematics (0.094).

This study found that there is a positive effect on academic achievement from test-based grade retention. As stated before, this study focused solely on the academic effects of retention and was consistent with the findings of other similar studies on the relationship of student achievement and statewide test-based retention policies (Lorence, 2014; Schwerdt, West, & Winters, 2017; Winters & Greene, 2012). The longitudinal multilevel modeling analysis is the best indication of the effect of retention since it controls for other factors that could be related to student performance. This type of analysis is something that many other studies on retention have not included, but long-term effects should always be considered in matters of educational policy (Brown, 2007). Since the findings of some other studies included that the positive

Table 16

Effects of Retention in ELA and Math Using Longitudinal Multilevel Modeling with School District Nesting

	Estimate	Std. Error	t-value	95% Conf. Interval	
				Lower Bound	Upper Bound
Grades 3-6					
ELA Z-scores					
Retained	0.018	0.036	0.502	-0.053	0.089
Math Z-scores					
Retained	0.063	0.036	1.725	-0.008	0.133
Grades 3-7					
ELA Z-scores					
Retained	0.032	0.032	1.009	-0.030	0.095
Math Z-scores					
Retained	0.065*	0.032	2.032	0.002	0.129
Grades 3-8					
ELA Z-scores					
Retained	0.113*	0.030	3.769	0.054	0.172
Math Z-scores					
Retained	0.094*	0.030	3.092	0.035	0.154

*Notes. * The mean difference is significant at the .05 level.*

academic effects of grade retention diminish as time passes from the time of retention (Allensworth & Nagaoka, 2010; Schwerdt, West, & Winters, 2017), I decided to look at what the model would show over shorter time periods. Table 16 provides a comparative look at how the effect of retention changes as years are added to the longitudinal multilevel analysis. That the effect of retention increases and becomes significant with additional years seems to imply that the relationship is not particularly strong as the additional years of data are necessary to reveal it as significant. However, it also implies that the effect does not simply fade out over years as some other studies have found. Additionally, the positive effects of retention were even greater when students with IEP or ELL exemptions were removed from the analysis.

Implications for Practice and Policy

It is my sincere hope that decision-makers at all levels of the educational process—parents, teachers, administrators, and policymakers—will benefit from what this study contributes to the body of research on the use of test-based grade retention, which will in turn benefit the students in their care. The decision to retain a student at any grade level is a difficult one and should never be handled in a cavalier manner. In my experience, the vast majority of educators view retention decisions in this manner, and my desire in undertaking this study was to provide additional light for those who would be facing such a determination. It is crucial to take into consideration the best available data and to make as well-informed a decision as possible with an understanding that students do not get to turn back time. In the decision-making process, the whole child should always be kept in view—meaning both academic and social/emotional outcomes. As I said before, the body of research literature on the subject of social and emotional effects of grade retention has been decidedly on the negative side. Logically, it would follow that there would need to be other clear reasons, such as definite academic benefits, to outweigh

the potential cons of socio-emotional damage, higher financial costs, higher dropout rates, and higher likelihood of discriminatory practices on the most vulnerable (Huddleston, 2014; Xia & Kirby, 2009; Yang, 2019).

While the results of this study show a positive academic effect for retention, there are some important limitations to keep these findings in perspective. The use of the released data from the Indiana Department of Education did not allow for an understanding of many additional factors. One of the principal considerations is that there is no way of knowing about the extent of the use of remediation services and other academic supports that might have accompanied the decision to retain. Duke, Moje, and Palincsar (2014) note that it is entirely unclear whether the short-term gains in achievement for students who were retained with their Grade 3 reading test policy were due to the retention or to the millions of dollars allocated for accompanying remediation and reform efforts. The inclusion of a nested level for the school corporation in the longitudinal model attempts to account for some of this variation, but there are differences between individual school buildings and individual teachers that cannot be accounted for with this dataset.

As Lorence (2014) suggested, "The degree of academic improvement possible after repeating a grade is likely linked to the reasons for the retention" (p.16). The grade of the retention and reasons for retaining may be very important. There certainly seems to be a difference between the efficacy of test-based and teacher-based retentions, and there are far more student-specific factors beyond even this distinction. This study cannot directly address reasons for retention and can only attest to the decision to retain at the end of Grade 3 based on a minimum proficiency reading assessment. Other studies have found less favorable results at

earlier or later grade levels (Hong & Raudenbush, 2005; Jacob & Lefgren, 2009; Manacorda, 2012).

The question remains of whether other remediation and intervention options are equally effective and more cost-efficient while also avoiding potential negative outcomes. The meta-analyses of Holmes (1989) as well as Xia and Kirby (2009) found studies that resulted in positive effects from retention included additional interventions. If retention costs an average of around $12,000 for an additional year of schooling, it is possible that funding other high-quality interventions could have similar or even superior benefits while costing taxpayers less money. The entire $12,000 could be saved in cases where additional intervention and remediation measures are being implemented concurrent with the retention. It is important to consider whether the benefits truly outweigh the costs—social/emotional damage for the students and their families as well the financial bottom line for schools.

The old saying about a definition of insanity being "doing the same thing over and over again and expecting different results" applies in many areas, and it has often been apropos with the practice of retention. Even if the specific retention policy does not require other interventions and reforms, the typical practices of schools can end up being significantly affected by the prospect of mandated retention. In view of the high-stakes nature of the IREAD-3 policy and the other accountability policies implemented in Indiana around the same time, there are almost certainly additional supports included in most if not all of the schools included in this study with many of them being put in place over the same time period covered in this dataset. These changes would also hold true in other recent test-based retention policy implementations. This paradigm shift from the days of just putting the student back into the same grade may be the most critical piece of the effectiveness of such policies. The impact of the cumulative effect of

these reforms are not only felt post-retention. These changes in practice to emphasize remediation and interventions for struggling students extend back into years prior to Grade 3 as schools seek to avoid results of Did Not Pass each year. It is probably best to view the positive effect of grade retention found in this study as an indication of the aggregate effect of all of the interventions associated with the IREAD-3 policy and not solely the practice of grade retention.

Finally, as Brown (2007) noted in his analysis of the Wisconsin test-based retention policy, it is extremely easy to lose sight of the real issue for individual students and make policies that focus on problems of the system. It can be exceedingly difficult to change the agenda of such policy reforms, so it is imperative to keep the best interests of individual students in focus along with any goals for systemic transformations. Educators must keep long-term effects, both positive and negative, in the spotlight when questions of policy are concerned in order to properly frame the perceived problem and provide viable alternatives to wrongheaded champions of reform. In the case of grade retention policies, there is more than one way to get things done, and finding low-cost and effective options is central to the work of school leaders.

Recommendations for Future Research

This study, to the extent of my knowledge and research, is the first to specifically analyze the effectiveness of the IREAD-3 retention policy in Indiana. There have been many similar studies conducted in other locations such as Chicago, Florida, and Texas, (Lorence, 2014; Schwerdt, West, & Winters, 2017; Winters & Greene, 2012; a summary of these studies is provided in Table 17), but this study was designed to utilize the state assessments and specifics of Indiana law and policy. While this study benefitted from its comprehensiveness with all student data available for the 2012 Grade 3 cohort, this study was still limited in its ability to

Table 17

Summary of Studies on Student Academic Achievement After the Decision to Retain in Large-scale Test-based Retention Policies in US Cities or States

Author(s)	Date	Data and Methods Used	Findings
Lorence	2014	Texas 1994-2002; Regression, Propensity Score Matching, Two-Level Hierarchical Linear Model; Same-grade comparison	Students who failed the Grade 3 reading test and were retained substantively outperformed their socially-promoted peers
Mariano & Martorell	2013	New York City 2004-2008; Regression Discontinuity; Same-grade comparison	Students retained in Grade 5 had a substantial, positive gain when compared to their socially-promoted peers which lasted for at least 2 years
Roderick & Nagaoka	2005	Chicago Public Schools 1997-2000; Regression Discontinuity; Three-Level Hierarchical Linear Model; Same-grade comparison	Students retained in Grade 3 had a small performance gain compared to socially-promoted peers that disappeared within 2 years; students retained in Grade 6 performed lower than their promoted peers
Schwerdt, West, & Winters	2017	Florida 2000-2014; Regression Discontinuity, Two-stage Least Squares Model; Same-grade and Same-age comparisons	Students retained in Grade 3 experienced substantial short-term gains compared to their socially-promoted peers; these effects fade out over the next 5 years on same-age comparison but remain significant on same-grade comparison
Winters & Greene	2012	Florida 2002-2009; Regression Discontinuity; Same-grade comparison	Effects of remediation policy (which includes retention, summer school, etc.) was substantially positive after 5 years though it dissipates over time

account for the some of the numerous possible factors with student achievement. For future studies, it would be helpful to be able to differentiate more on the actual decision to retain. In this dataset, it was impossible to distinguish between the mixture of IREAD-3 retentions and other retentions as reasons for retention were not reported and, hence, not available. A smaller and more detailed dataset could provide additional insights into the effects of additional factors on the overall academic outcomes observed. There would also be an opportunity to use qualitative methods to follow up with individual students, parents, and educators to tease out differences in experiences and ancillary factors in retention determinations as well as practices of schools and districts when retention decisions are made.

Same-grade vs. same age comparisons performed on the same dataset could be possible on an assessment that is vertically aligned in its scoring, which is what Schwerdt, West, and Winters (2017) were able to do in their analysis of Florida's policy. I believe same-grade comparisons are more appropriate to the retention question, but there is some disagreement on this point. Logically, same-age comparisons would be less likely to show positive results for retention, a difference which Allen, Chen, Willson, and Hughes (2009) point out. While I chose to use the statewide dataset for this study, using other measures of student performance such as NWEA MAP Growth which are specifically aligned from year-to-year could provide different results than were seen using ISTEP+. That also brings up the question of the validity of ISTEP+ as an assessment of academic progress to begin with (Grissmer, Ober, & Beekman, 2014). However, for this study on a specific Indiana policy, using the same measure that is used for accountability purposes in Indiana makes sense. Its use also allowed for a comprehensive sample to analyze the effectiveness of this statewide policy. Nonetheless, it was unfortunate to have to deal with a change in the assessment itself in the middle of the years examined in this

study, so future studies may be able to avoid any concerns in that area by using a consistent test measure rather than relying on an after-the-fact standardized score from a scale score.

Another factor specific to Indiana and the timeframe covered in this study is that many additional teacher and school accountability measures increased in the state during the years examined. These could have contributed to increased attention on failing students and changes in practices at the classroom, school, and district levels. However, it seems likely that these would have benefited the socially promoted students and retained students alike. Additionally, as Hout, Elliott, and Freuh (2012) found that while the threat of high-stakes tests did not carry over into an increase in learning, these other measures may or may not have made a difference in student achievement. Regardless, future studies on grade retention policies might focus on these supplemental supports as they relate to grade retention and subsequent academic outcomes for individual students.

Concluding Remarks

This study demonstrates that the potential exists for positive academic outcomes from grade retention. As the IREAD-3 policy provides for, it is incumbent upon groups of adults who know and care about the child—teams of educators in collaboration with parents—to carefully consider matters such as grade retention and, in good faith, make the best choice they can with the best information available. While researchers continue to study grade retention resulting in a mixture of outcomes, the values of many other educational practices are clearly understood. Practices such as early identification of students with academic struggles and high-quality intervention strategies—whether through RtI, extended-day, or extended-year programs—should be the focus for educators and policymakers over concerns with retention practices. For school leaders who are making a decision regarding grade retention, it is highly recommended that they

carefully consider all aspects of a child's development with an emphasis on objective measures over informal observational data such as parent or teacher opinions of their classroom performance and behavior. Retention could be considered in very specific circumstances due to lack of results from interventions and occurrence of life events for an individual student, but only after extensive intervention strategies have been employed. With an issue as complex and as potentially far-reaching in impact as the decision to retain a student, it is always difficult to dissect all the factors involved and distinguish which ones should guide the decision-making. In this arena, as in so many others where individual students are involved, there is far more gray than many people are comfortable with or certainly would prefer. We must learn to become comfortable with not definitively knowing and focus on striving to see students as the individuals that they are and on working together to help ensure a bright future.

References

Adams, C. J., Robelen, E. W., & Shah, N. (2012). Data Show Retention Disparities. *Education Week*, *31*(23), 1-18.

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades*. Cambridge University Press.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, *31*(4), 480-499.

Allensworth, E. M., & Nagaoka, J. (2010). Issues in studying the effects of retaining students with high-stakes promotion tests: Findings from Chicago. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schooling, and human development* (pp. 327-341). New York, NY: Routledge.

Ashby, C. M. (2010). *K-12 education: Many challenges arise in educating students who change schools frequently. Report to congressional requesters.* GAO-11-40. Washington, D.C.: US Government Accountability Office.

Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. New York, NY: Routledge.

Beebe-Frankenberger, M., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Sorting second-grade students: Differentiating those retained from those promoted. *Journal of Educational Psychology*, *96*(2), 204–215. https://doi.org/10.1037/0022-0663.96.2.204

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* New York, NY: Basic Books.

Bolon, C. (2000). School-based standard testing. *Education Policy Analysis Archives*, *8*(23), 1-43.

Booher-Jennings, J. (2008). Learning to label: Socialisation, gender, and the hidden curriculum of high-stakes. *Journal of Sociology of Education*, *29*(2), 149–160. https://doi.org/10.1080/01425690701837513

Bowman, L. J. (2005). Grade retention: Is it a help or hindrance to student academic success? *Preventing School Failure.*, *49*(3), 42–46. https://doi.org/10.3200/PSFL.49.3.42-46

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513-531.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by design and nature.* Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1994). Ecological models of human development. *International Encyclopedia of Education, Vol. 3,* (2nd Ed.). Oxford, UK: Elsevier.

Bronfenbrenner, U, & Ceci, S. J. (1994). Nature-nurture reconceptualization in developmental perspective: A bioecological model. Psychological Review, 101(4), 568-586.

Bronfenbrenner, U. (1995). Developmental ecology through space and time: A future perspective. In P. Moen, G. H. Elder, Jr., & K. Lüscher (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 619–647). American Psychological Association.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 993-1028). Hoboken, NJ: John Wiley.

Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and operational models. In S. L. Friedman & T. D. Wachs (Eds.), Measuring Environment Across the Life Span: Emerging Methods and Concepts (pp. 3-28). Washington, DC: American Psychological Association Press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of Child Psychology, Vol. 1, Theoretical Models of Child Development* (6th ed., pp. 793-828). New York, NY: Wiley.

Brown, C. P. (2007). Examining the streams of a retention policy to understand the politics of high- stakes reform. *Education Policy Analysis Archives, 15*(9). Retrieved from http://epaa.asu.edu/epaa/v15n9/

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education, 57*(3), 190-213.

Burger, J. M., & Krueger, M. (2003). A balanced approach to high-stakes achievement testing: An analysis of the literature with policy implications, *International Electronic Journal for Leadership in Learning, 7*(4).

Burkam, D. T., Lee, V. E., & Dwyer, J. (2009). *School mobility in the early elementary grades: Frequency and impact from nationally-representative data.* Paper prepared for the Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods, Washington, DC.

Burkam, D. T, LoGerfo, L., Ready, D., & Lee, V. E. (2007). The differentials effects of repeating kindergarten. *Journal of Education for Student Placed At Risk, 12*(2), 103-136.

Butler, S. R., Marsh, H. W., Sheppard M. J., & Sheppard J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology*, *77*(3), 349–361.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67-90.

Cannon, J. S., & Lipscomb, S. (2011). *Early grade retention and student success: Evidence from Los Angeles*. Public Policy Instit. of CA.

Cockx, B., Picchio, M., & Baert, S. (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics*, *34*(3), 403-424.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2d ed. Hillsdale, N.J.: Lawrence Erlbaum.

Cohen, J., Loeb, S., Miller, L. C., & Wyckoff, J. H. (2020). Policy implementation, principal agency, and strategic action: Improving teaching effectiveness in New York City middle schools. *Educational Evaluation and Policy Analysis*, *42*(1), 134-160.

Comber, B. (2011). Changing literacies, changing populations, changing places--English teachers' work in an age of rampant standardisation. *English Teaching: Practice and Critique*, *10*(4), 5-22.

Creswell, J. W. (2009). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Crosnoe, R. (2006). The connection between academic failure and adolescent drinking in secondary school. *Sociology of Education*, *79*(1), 44-60.

Dawson, P. (1998, June). A primer on student grade retention: What the research says. *NASP Communique*, *28*(8).

Denton, D. R. (2001, January). *Finding alternatives to failure: Can states end social promotion and reduce retention rates?* Atlanta, GA: Southern Regional Education Board.

DiMaria,Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago. *American Educational Research Journal*, *44*(3), 594–629. https://doi.org/10.3102/0002831207306767

Duke, N. K., Moje, E. B., & Palincsar, A. S. (2014). Three IRA literacy research panel members comment on third grade retention laws. *Michigan Reading Journal*, *47*(1), 11.

Eren, O., Lovenheim, M. F., & Mocan, N. H. (2018). *The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy* (No. w25384). National Bureau of Economic Research.

Fanguy, J., & Mathis, R. (2012). Psychosocial fallout from grade retention: Implication for educators. *Delta Journal of Education*, *2*(2), 69-82.

Feuer M. L. (1992). *Testing in American schools: Asking the right questions.* Washington, DC: U. S. Congress, Office of Technology Assessment.

Fiester, L. (2013). *Early warning confirmed: A research update on third-grade reading.* Baltimore, MD*:* The Annie E. Casey Foundation. Retrieved from https://www.aecf.org/m/resourcedoc/AECF-EarlyWarningConfirmed-2013.pdf

Fine, J. G., & Davis, J. M. (2003). Grade retention and enrollment in post-secondary education. *Journal of School Psychology*, *41*(6), 401-411.

Fowler, F. C. (2013). *Policy studies for educational leaders* (4th ed.). Boston, MA: Pearson.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1966). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17.

Frey, N. (2005). Retention, social promotion, and academic redshirting: What do we know and need to know? *Remedial and Special Education*, *26*(6), 332–346.

Geiser, S., & Santelices, M.V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* Berkeley: Center for Studies in Higher Education, University of California, Berkeley. Retrieved from https://escholarship.org/uc/item/7306z0zf

Gleason, K. A., Kwok, O., & Hughes, J. N. (2007). The short-term effect of grade retention on peer relations and academic performance of at-risk first graders. *The Elementary School Journal*, *107*(4), 327–340.

Glossary of Education Reform. (n.d.). Retrieved September 21, 2019, from https://www.edglossary.org/standardized-test/

Goddard, R. D., Skrla, L., & Salloum, S. J. (2017). The role of collective efficacy in closing student achievement gaps: A mixed methods study of school leadership for excellence and equity. *Journal of Education for Students Placed at Risk (JESPAR)*, *22*(4), 220-236.

Gottfredson, D. C., Fink, C. M., & Graham, N. (1994). Grade retention and problem behavior. *American Educational Research Journal*, *31*(4), 761-784.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test- based promotion policy. *Education Finance and Policy, 2*, 319-340. http://dx.doi.org/10.1162/edfp.2007.2.4.319

Grissmer, D. W., Ober, D. R., & Beekman, J. A. (2014). Focusing on short-term achievement gains fails to produce long-term gains. *Education Policy Analysis Archives, 22*(5), 1-36.

Hall, W. J., & Chapman, M. V. (2018). The role of school context in implementing a statewide anti-bullying policy and protecting students. *Educational Policy, 32*(4), 507-539.

Hamilton, L., Stecher, B., Marsh, J., McCombs, J., Robyn, A., Russell, J., … Barney, H. (2007). Standards-based accountability under no child left behind: Experiences of teachers and administrators in three states. In *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States*. https://doi.org/10.7249/mg589

Hattie, J. A. (2012). Visible learning: A synthesis of 800+ meta-analyses on achievement. Abingdon, UK: Routledge.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS*. (2nd ed.) Routledge.

Henry, G. T., Rickman, D. K., Fortner, C. K., & Henrick, C. C. (2005). *Report of the findings from Georgia's third grade retention policy.* Atlanta, GA: Andrew Young School of Public Policy.

Hernandez-Tutop, J. (2012). Social promotion or grade repetition: What's best for the 21st Century student? [Online Submission]. Retrieved from http://files.eric.ed.gov/fulltext/ED532287.pdf

Heubert, J. and Hauser, R.M. (1999). *High stakes testing for tracking, promotion, and graduation*. Washington, D.C.: National Academy Press.

Hinojosa, M. S., Hinojosa, R., Bright, M., & Nguyen, J. (2019). Adverse childhood experiences and grade retention in a national sample of US children. *Sociological Inquiry*, *89*(3), 401-426.

Holmes, C. T., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*(2), 225–236.

Holmes, C.T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L.A. Shepard & M.L. Smith (Eds.), *Flunking Grades: Research and Policies on Retention* (pp. 16–33). London: Falmer Press.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*(3), 205-224.

Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, *29*(4), 239-261.

Hong, G., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology*, *44*(2), 407.

Honig, M. I. (2006). Complexity and policy implementation: Challenges and opportunities for the field. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1 – 23). Albany, NY: SUNY Press.

Hout, M., Elliott, S., & Frueh, S. (2012). Do high-stakes tests improve learning? *Issues in Science and Technology*, 33–38. Retrieved from http://web.b.ebscohost.com.libproxy.noctrl.edu/ehost/pdfviewer/pdfviewer?sid=a786e66a-26b3-4cf9-99ce-b9a84e4b664f%40sessionmgr110&vid=1&hid=128

Huddleston, A. P. (2014). achievement at whose expense? A literature review of test-based grade retention policies in U.S. schools. *Education Policy Analysis Archives*, *22*(18). https://doi.org/10.14507/epaa.v22n18.2014

Huddleston, A. P. (2015). "Making the Difficult Choice": Understanding Georgia's Test-Based Grade Retention Policy in Reading. *Education Policy Analysis Archives*, *23*(51).

Hughes, J. N., Cao, Q., West, S. G., Smith, P. A., & Cerda, C. (2017). Effect of retention in elementary grades on dropping out of school early. *Journal of School Psychology*, *65*, 11-27.

Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2018). Effect of early grade retention on school completion: A prospective study. *Journal of educational psychology*, *110*(7), 974.

Hwang, S. H. J., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness*, *11*(4), 559–587. https://doi.org/10.1080/19345747.2018.1496500

Indiana Department of Education (July 13, 2018). *IREAD-3 Policy and Guidance*. https://www.doe.in.gov/assessment/iread-3-policy-and-guidance

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, *1*(3), 33–58. https://doi.org/10.1257/app.1.3.33

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, *37*(3), 243-272.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420–437. https://doi.org/10.1177/017084068800900203

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools*, *39*(4), 441–457. https://doi.org/10.1002/pits.10046

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly*, *22*(3), 314–339. https://doi.org/10.1037/1045-3830.22.3.314

Jimmerson, S. R., & Renshaw, T. L. (2012). Retention and social promotion. *Principal Leadership*, 12–16. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.348.6725&rep=rep1&type=pdf

Katz, L. G. (2000). *Academic redshirting and young children*. ERIC Digest. Retrieved from https://files.eric.ed.gov/fulltext/ED447951.pdf

Keyes, C. H. (1911). *Progress Through the Grades of City Schools: A Study of Acceleration and Arrest* (No. 42). Teachers College, Columbia University. Retrieved from https://books.google.com/books?hl=en&lr=&id=R40VAAAAIAAJ&oi=fnd&pg=PA1&dq=Keyes,+C.+H.+(1911).+Progress+through+the+grades+of+city+schools+(Book+Review).+Journal+of+Education,+74,+269-270.&ots=nKsqY88Kl1&sig=hjmcxXe7hFJOSBKT-k0X2ttYA00#v=onepage&q&f=false

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools.* Portsmouth, NH: Heineman.

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Cambridge, MA: Harvard University Press.

Larsen, D. E., & Akmal, T. T. (2007). *Making decisions in the dark: Disconnects between retention research and middle-level practice*. https://doi.org/10.1177/0192636506298832

LiCalsi, C., Ozek, U., & Figlio, D. (2019). The Uneven Implementation of Universal School Policies: Maternal Education and Florida's Mandatory Grade Retention Policy. *Education Finance and Policy, 14*(3), 383-413.

Lorence, J. (2014). Third-grade retention and reading achievement in Texas: A nine year panel study. *Social Science Research, 48*, 1-19. http://dx.doi.org/10.1016/j.ssresearch.2014.05.001

Lynch, M. (2014). The true costs of social promotion and retention. *International Journal of Progressive Education, 10*(3), 6-17.

Manacorda, M. (2012). The cost of grade retention. *The Review of Economics and Statistics*, 94. Retrieved from https://www.jstor.org/stable/pdf/23262090.pdf?casa_token=7bJSISXSonwAAAAA:vyhHql-arwuG6XWlhf-da5i9AHodir1fHvCZRJnUo91SUDhmwq0Q5vmfojgK_yS70z47TPEYMws52SgT4QSEYnnHlZ4kxfseAVY_enO_D4HZQONHbOw

Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis, 35*(1), 96-117.

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology*, *109*(3), 425.

McCarty, C. A., Mason, W. A., Kosterman, R., Hawkins, J. D., Lengua, L. J., & McCauley, E. (2008). Adolescent school failure predicts later depression among girls. *Journal of Adolescent Health, 43*(2), 180-187.

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, *37*(3), 273-298.

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly, 15*(4), 351-377.

McMahon, T. (2018). Despite our best intention: Students relate how social promotion hurt them and what changes they believe will help them. *Interchange, 49*(4), 499-519.

Meisels, S. J., & Liaw, F. R. (1993). Failure in grade: Do retained students catch up? *Journal of Educational Research, 87*(2), 69-77.

Mehana, M., & Reynolds, A. J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review, 26*(1), 93-119.

Mordica, J. (2006). *Third grade students not meeting standards in reading: A longitudinal study*. (Research Brief No. 02). Atlanta, GA: Georgia Department of Education.

National Association of Psychologists (2011). *Grade retention and social promotion*.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: Author.

National Research Council and Institute of Medicine. (2010). *Student mobility: Exploring the impact of frequent moves on achievement [Summary of a workshop]* (A. Beatty, Ed.). Washington, DC: National Academies Press.

Ou, S. R., & Reynolds, A. J. (2008). Predictors of educational attainment in the Chicago longitudinal study. *School Psychology Quarterly*, *23*(2), 199-229.

Owen, S. A., & Ranick, D. L. (1977). The Greensville program: A commonsense approach to basics. *Phi Delta Kappan*, *58*, 531-533, 539.

Owings, W. A., & Kaplan, L. S. (2001). Standards, retention, and social promotion. *NASSP Bulletin*, *85*(629), 57–66. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/019263650108562906?casa_token=vulCZfMzPdUAAAAA:FhioX56DJOkew-CsKnVQu528Zg-Jg7r1ivSdPMB9OVUb5CGg4sA6S4T0BJJQosu-5Ekn62Q4pflD

Peterson, L. S., & Hughes, J. N. (2011). The differences between retained and promoted children in educational services received. *Psychology in the Schools*, *48*(2), 156–165. https://doi.org/10.1002/pits.20534

Raffaele Mendez, L. M., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering school progression through delayed entry or kindergarten retention: Propensity score analysis of long-term outcomes. *Journal of Educational Research*, *108*(3), 186–203. https://doi.org/10.1080/00220671.2013.867474

Renaud, G. A. (2013). Grade retention: Elementary teacher's perceptions for students with and without disabilities. *SAGE Open*, *3*(2), 1–16. https://doi.org/10.1177/2158244013486993

Reschly, A. L., & Christenson, S. L. (2013). Grade retention: historical perspectives and new research. *Journal of School Psychology*, *51*(3), 319-322.

Reynolds, A. J., Chen, C., & Herbers, J. (2009, June). *School mobility and educational success: A research synthesis and evidence on prevention.* Paper presented at the National Academies Workshop on the Impact of Mobility and Change on the Lives of Young Children, Schools, and Neighborhoods, Washington, DC.

Richman, J. M., Bowen, G. L., & Woolley, M. E. (2004). School failure: An eco-interactional developmental perspective. In M. W. Fraser (Ed.), *Risk and resilience in childhood: An ecological perspective*, (pp. 133-160). Washington, DC: NASW press.

Roach, M., & Kloosterman, P. (2014). 2013 NAEP: How Does Indiana Compare? *Center for Evaluation and Education Policy Education Policy Briefs*, *12*(1), 1-12.

Robertson, R. M. (2021). To Retain or Not Retain: A Review of Literature Related to Kindergarten Retention. *Online Submission*. Retrieved from https://files.eric.ed.gov/fulltext/ED611006.pdf

Robinson, V. M. J., Lloyd, C.A., & Rowe, K.J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635-674.

Roderick, M. R. (1993). *The path to dropping out: Evidence for intervention*. Auburn House.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis, 27*, 309-340. http://dx.doi.org/10.3102/01623737027004309

Rose, S., & Schimke, K. (2012). Third grade literacy policies: Identification, intervention, retention. *Education Commission of the States (NJ3)*. Retrieved from https://files.eric.ed.gov/fulltext/ED535949.pdf

Rumberger, R. W. (2003). The causes and consequences of student mobility. *Journal of Negro Education*, 72(1), 6-21.

Schnurr, B. L., Kundert, D. K., & Nickerson, A. B. (2009). Grade retention: Current decision-making practices and involvement of school psychologists working in public schools. *Psychology in the Schools, 46*(5), 410-419.

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*. Retrieved from http://www.nber.org/papers/w21509

Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention.* London, England: Falmer Press.

Silberglitt, B., Jimerson, S. R., Burns, M. K., & Appleton, J. J. (2006). Does the timing of grade retention make a difference? Examining the effects of early versus later retention. *School Psychology Review*, *35*(1), 134–141. Retrieved from https://www.essr.net/~jafundo/mestrado_material_itgjkhnld/SP/Sucesso e insucesso/Timing da retenção.pdf

Skiba, R., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores. Children left behind policy briefs supplementary analysis 2-D.* Bloomington, IN: Indiana University Center for Evaluation and Education Policy.

Smalls, U. (1997). *Reacting in the best interest of our students.* Retrieved from https://files.eric.ed.gov/fulltext/ED415980.pdf

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334-344.

Spillane, J. P. (2005). Standards deviation: How schools misunderstand education policy. *CPRE Policy Briefs.* Retrieved from http://repository.upenn.edu/cpre_policybriefs/31

Starr, A. (2019, July 13). *States are ratcheting up reading expectations for 3rd-graders.* [Radio Broadcast]. NPR. https://www.npr.org/2019/07/13/741156019/states-are-ratcheting-up-reading-expectations-for-3rd-graders

Stearns, E., Moller, S., Blau, J., & Potochnick, S. (2007). Staying back and dropping out: The relationship between grade retention and school dropout. *Sociology of Education, 80*(3), 210-240.

Stubbs, V. F. (2013). Retention as a state policy mandate: IREAD in Indiana. Retrieved from http://ezphost.dur.ac.uk/login?url=https://search.proquest.com/docview/1430500710?accountid=14533%0Ahttp://openurl.ac.uk/ukfed:dur.ac.uk?genre=dissertations+%26+theses&issn=&title=Retention+as+a+state+policy+mandate%3A+IREAD+in+Indiana&volume=&issue=&date

Third-Grade Reading Legislation (NCSL). (2019). Retrieved from: http://www.ncsl.org/research/education/third-grade-reading-legislation.aspx

Tudge, J. R., Mokrova, I., Hatfield, B. E., & Karnik, R. B. (2009). Uses and misuses of Bronfenbrenner's bioecological theory of human development. *Journal of Family Theory & Review, 1*(4), 198-210.

U.S. Census Bureau. (2017). *2017 Public Elementary-Secondary Education Finance Data*. Retrieved from https://www.census.gov/data/tables/2017/econ/school-finances/secondary-education-finance.html

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). The effects of early grade retention: Effect modification by prior achievement and age. *Journal of School Psychology, 54*, 77-93.

Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys, 35*(2), 408-451.

Welsh, R. O. (2017). School hopscotch: A comprehensive review of K–12 student mobility in the United States. *Review of Educational Research, 87*(3), 475-511.

Winters, M. A. (2019). *The effect of Florida's test-based promotion policy on student performance prior to the retention decision*. Retrieved from http://sites.bu.edu/marcuswinters/files/2018/10/Effect-of-Test-Based-Promotion-Prior-to-Retention-Decision-Working-paper.pdf

Winters, M. A., & Greene, J. P. (2006). Getting ahead by staying behind: An evaluation of Florida's program to end social promotion. *Education Next, 6*(2), 65-69.

Winters, M. A., & Greene, J. P. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy, 7*(3), 305-330. http://dx.doi.org/10.1162/EDFP_a_00069

Wirtz, W. (1977). *On further examination: Report of the advisory panel on the Scholastic Aptitude Test score decline.* New York, NY: College Entrance Examination Board.

Worthen, B.R., & Spandel, V. (1991). Putting the standardized test debate in perspective. *Educational Leadership, 48*(5), 65-69.

Xia, N., & Kirby, S. N. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes.* (Technical Report No. 678). Santa Monica, CA: Retrieved from: http://www.rand.org/pubs/technical_reports/TR678/

Yang, M. (2019). *The effect of grade retention on academic and social-behavioral outcomes for students with disabilities in elementary grades* (Doctoral dissertation).

Appendix A

EXHIBIT A

For the Data Sharing Agreement entered into between the Indiana Department of Education and Vince Edwards

I. Scope of Work

Background and Purpose

I am studying grade retention, specifically test-based policies such as the one pursuant to Public Law 109 from 2010 that initiated the development of IREAD-3 and subsequent requirements in Indiana. Since it was first given in spring 2012, we would now have data from ISTEP+ through that cohorts grade 8 assessment. I want to examine the effect of grade retention as an intervention for those students from an academic achievement perspective as measured by ISTEP+.

Research Questions and Design

I have chosen to use a non-experimental quantitative approach to study the effects of grade retention due to the IREAD-3 policy on student achievement levels. This will be a longitudinal study for students in the first cohort to take IREAD-3 under the policy in 2012 from third through eighth grade and will be done using something of a Control-Group Interrupted Time-Series Design. I plan to compare students within a range of the pass cut score (above and below) which would be determined by statistical analysis of the scores from across the state. I want to see if they were, in fact, retained in 3rd grade according to the policy and how they performed on subsequent ISTEP+ tests through 8th grade. While this study contributes to an understanding of the effectiveness of test-based retention policies in general, the specific research questions of this study relate to effects on student academic achievement of the IREAD-3 policy in the state of Indiana. This study has one primary research question and three ancillary questions that will directly inform the primary question.

> Primary Research Question: Is there a positive relationship between grade retention in accordance with the IREAD-3 policy in Indiana and subsequent student achievement outcomes?
>
> Subquestion 1: Is there a positive relationship between grade retention and student achievement one year after the decision to retain?
>
> Subquestion 2: Is there a positive relationship between grade retention and student achievement two years after the decision to retain?
>
> Subquestion 3: Is there a positive relationship between grade retention and student achievement six years after the decision to retain?

The choice to use six years after the decision to retain arises from the current grade level and the data available. Six years after the decision to retain would place the student in 8th grade which is on the cusp of entering high school and is also the last year of continuous achievement data from the ISTEP+ assessment.

Reporting and Dissemination Plan

II. Description of Data Being Requested

Description of Data Needs

I am seeking to get IREAD-3 scale scores and ISTEP+ scale scores (Math and E/LA) in grades 3-8 for all students in Indiana since 2012 when IREAD-3 was first administered. I am requesting de-identified data and would like to have the following student demographic data included with the test results in order to control for various factors in the statistical analysis:

1. School district
2. School
3. Student Alternate ID
4. Grade level
5. Days in attendance at school (pulled from AT report)
6. Birthdate
7. Gender
8. Ethnicity
9. Free/reduced lunch status
10. IREAD-3 IEP Exemption status
11. IREAD-3 EL Exemption status
12. IREAD-3 scores
13. ISTEP+ Math Scale scores
14. ISTEP+ E/LA Scale scores

Data Period

Spring 2012 – Summer 2018

Appendix B

States that require retention of third graders who fail a minimum reading proficiency test:

Arizona
Arkansas
California
Connecticut
Delaware
District of Columbia
Florida
Georgia
Indiana
Iowa
Michigan
Mississippi
Missouri
Nevada
North Carolina
Ohio
South Carolina
Tennessee
Washington

States that allow retention of third graders who fail a minimum reading proficiency test (but do not require retention):

Alabama
Alaska
Colorado
Maine
Maryland
Minnesota
New Jersey
New Mexico
Oklahoma
Texas
West Virginia

Source: (NCSL, 2019)

Appendix C

Indiana DEPARTMENT OF EDUCATION

Dr. Jennifer McCormick
Superintendent of Public Instruction
Working Together for Student Success

IREAD-3: Frequently Asked Questions

<table>
<tr><th>Question</th><th>Answer</th></tr>
<tr><td>1. Which students are required to participate in IREAD-3?</td><td>Third grade students enrolled in accredited public and nonpublic schools who participate in ISTEP+ must participate in IREAD-3.
<table>
<tr><th>Students</th><th>Participate in Assessment</th></tr>
<tr><td>Students with a Section 504 Plan</td><td>Yes</td></tr>
<tr><td>Students with an IEP</td><td>Yes</td></tr>
<tr><td>Students with an ILP</td><td>Yes</td></tr>
<tr><td>Students who have been retained twice</td><td>Yes</td></tr>
<tr><td>Students who are eligible for the alternate assessment</td><td>No</td></tr>
</table></td></tr>
<tr><td>2. Must students who have been in the U.S. for less than one year and who will take WIDA-ACCESS also participate in IREAD-3?</td><td>Yes.</td></tr>
<tr><td>3. What is the expectation for schools whose Spring Break is scheduled for the same week as the IREAD-3 assessment?</td><td>The 2017-2018 Assessment Date Change Request form is located in the program manual on our website (http://www.doe.in.gov/assessment). A school or corporation scheduled to be on Spring Break during the IREAD-3 assessment window must fax their completed form to the Office of Student Assessment by the deadline on the Date Change Request form.</td></tr>
<tr><td>4. Will IREAD-3 be administered in addition to ISTEP+ at grade three?</td><td>Yes. Separate from ISTEP+, IREAD-3 is the assessment Indiana uses to measure reading skills as part of the implementation of House Enrolled Act 1367 (also known as PL 109 in 2010).</td></tr>
<tr><td>5. Is IREAD-3 a section of ISTEP+, or does it replace the reading portion?</td><td>Neither, as IREAD-3 is a separate assessment. It does not replace any portion of the existing ISTEP+. Grade 3 students must participate in both ISTEP+ and IREAD-3.</td></tr>
<tr><td>6. Does IREAD-3 consist of multiple-choice items only?</td><td>Yes. Located on the IDOE web site, the IREAD-3 Item Sampler features the different types of multiple-choice questions that could appear on IREAD-3.</td></tr>
</table>

7. Has the state decided what defines "reading level"? Lexiles for example?	The IREAD-3 assessment measures foundational reading skills through grade three. As with all of our state assessments, a committee of Indiana educators recommended a passing or "cut" score to the State Board of Education; the State Board approved the recommended cut score at their August 2011 meeting. Meeting expectations for "reading level" were determined by the IREAD-3 cut score.
8. What IREAD-3 resources are available for parents and teachers?	There are several documents in the "Additional Resources" section of the IREAD-3 webpage: http://www.doe.in.gov/assessment/iread-3
9. Will IREAD-3 follow other state assessments with allowable accommodations that must be addressed in the IEP, Section 504 Plan, and/or an ILP (EL plan)?	As with all of our state assessments, the IREAD-3 test administration includes accommodations. See Appendix C of the *2016-17 Indiana Assessment Program Manual* for specific information: http://www.doe.in.gov/assessment
10. Are schools required to administer the IREAD-3 practice test to students before they take the IREAD-3 operational assessment? What if a student is absent when the practice test is administered?	Yes, schools are required to administer the IREAD-3 Practice Test to students ***before*** administering the IREAD-3 assessment. If a student is absent when the practice test is administered and returns to school during the test window, the school ***must*** administer the practice test to the student before administering the operational assessment to the student.
11. What types of reports are made available to parents and schools?	**Spring & Summer:** For each student who participates in IREAD-3, schools receive an *Individual Student Report* (ISR) and a label for the student's cumulative file. Like ISTEP+, student performance on IREAD-3 is reported by standard; the ISR lists the standards assessed and the student's percent correct for each standard. Schools provide parents with a copy of their child's *Individual Student Report*. In addition, the following reports are available for permission-level access by individual schools and corporations: For **administrators**: • Data file • Proficiency Roster • Individual Student Reports For **teachers**: • Proficiency Roster • Individual Student Reports
12. How can a corporation add and/or correct STN information in its data file?	Any updates to a student's STN should be processed through the STN Help Desk (1-800-527-4931). Updates will not impact the corporation's data file.

13. What score is needed to pass IREAD-3?	A student must achieve a scale score of 446 to pass IREAD-3. The list of IREAD-3 *Pass* and *Did Not Pass* Performance Level Descriptors contains more information (refer to the "Additional Resources" section of the IREAD-3 website: http://www.doe.in.gov/assessment/iread-3)
14. Is there a rescore process for IREAD-3?	No. A rescore process is offered only for those assessments that contain open-ended items (i.e., ISTEP+, ECAs); IREAD-3 consists of multiple-choice questions *only*, so no rescore process is available.
15. What happens in the event that a student's ISTEP+ ELA results differ from IREAD-3 results?	While IREAD-3 and ISTEP+ ELA both measure reading comprehension skills, they differ in that ISTEP+ ELA also assesses language arts standards; IREAD-3 assesses foundational reading skills *only*. Because the assessments are not identical, some differences in results should be expected.
16. How will the final results of IREAD-3 be calculated?	A school's percent passing IREAD-3 will be calculated after the spring (preliminary) and summer administrations (final).
17. Which students should receive IREAD-3 remediation?	Under State Board rule, schools must provide additional intervention(s) for students who do not pass IREAD-3, ***including students who receive a Good Cause Exemption***. Also, students who do not have the opportunity to participate in the Spring 2017 IREAD-3 assessment are ***eligible for intervention***, just as those students who did not pass the assessment.
19. What are the consequences for students who do not pass IREAD-3, who do not participate in either the spring or summer administration, or who do not meet any of the eligibility criteria for a Good Cause Exemption?	In accordance with HEA 1367 (also known as PL 109 in 2010), students who do not pass the Grade 3 reading assessment (IREAD-3) will: • continue to receive instruction in Grade 3 Reading, and • fully participate in the IREAD-3 assessment. Schools are responsible for designing a program that best meets the learning needs of students and to organize students for instruction. The Indiana State Board of Education approved guidance in Spring 2017 that provides schools flexibility to officially report grade-level designations for students who do not have a passing IREAD-3 score.
20. Would a student who does not pass IREAD-3, but whose school officially reports him or her as a Grade 4 student, take the Grade 3 or Grade 4 ISTEP+ in the spring?	Students must participate in the ISTEP+ that corresponds with the grade level in which they are enrolled/officially reported. A retest opportunity for IREAD-3 must also be provided.
21. If a student passes IREAD-3 and is retained for other reasons, will he/she need to take IREAD-3 the following spring?	No. Once a student achieves a passing score on IREAD-3, he/she will not need to retake the assessment.

<table>
<tr>
<td>22. What are the Good Cause Exemptions and which students are eligible?</td>
<td>
The Good Cause Exemptions include:

Students who have previously been retained two times prior to fourth grade.
Students with disabilities whose Case Conference Committees have determined that promotion is appropriate.
English Learner students whose Individual Learning Plan (ILP) Committees have determined that promotion is appropriate.

<table>
<tr><th>Students</th><th>Participate in the Assessment</th><th>Eligibility for Good Cause Exemption</th></tr>
<tr><td>Students with a Section 504 Plan</td><td>Yes</td><td>No</td></tr>
<tr><td>Students with an IEP (SWDs)</td><td>Yes</td><td>Determined by Case Conference Committee</td></tr>
<tr><td>Students with an ILP (ELs)</td><td>Yes</td><td>Determined by ILP Committee</td></tr>
<tr><td>Students who have been retained twice</td><td>Yes</td><td>Yes</td></tr>
<tr><td>Students who are eligible for the alternate assessment.</td><td>No</td><td>N/A</td></tr>
</table>
</td>
</tr>
<tr>
<td>23. Must students who have an IEP or ILP participate in the summer retest?</td>
<td>Yes, unless the student has received a Good Cause Exemption. The Case Conference Committee or ILP Committee makes this decision.</td>
</tr>
<tr>
<td>24. Suppose that a student did not pass IREAD-3 in spring or summer. If the student did not receive an IEP until after the spring (or summer) test window, would he become eligible for a Good Cause Exemption?</td>
<td>Yes. Once a student is determined eligible for special education services, the student becomes eligible for a Good Cause Exemption. The student's Case Conference Committee must meet to review the student's test results and other evidence and determine whether to grant an exemption.</td>
</tr>
<tr>
<td>25. Who is responsible for granting and/or reporting a Good Cause Exemption for a student who transfers to another school?</td>
<td>The Case Conference Committee or ILP Committee is responsible for meeting soon after the school receives the student's IREAD-3 results to determine whether to grant the student a Good Cause Exemption. If an exemption is granted to the student, his or her school should report the student's exemption as soon as possible via the DOE-IR collection in the STN Application Center.</td>
</tr>
</table>

26. Which students must participate in the summer retest?	The IREAD-3 Summer Retest is for Grade 3 students who: • did not pass IREAD-3 in March and have since been provided reading remediation/intervention • were absent during the March test window • moved into Indiana *after* the March test window and *before* the end of the school year
27. Must students who were homeschooled or who attended an out-of-state school pass IREAD-3 before they can be enrolled as 4th graders in an accredited Indiana School?	No. Students who are homeschooled do not participate in any state assessments such as IREAD-3. The receiving school administer local assessments to determine the appropriate grade-level placement for any new student. Students who were promoted to 4th grade by their teacher at an out-of-state school would not take IREAD-3. **Only 3rd graders enrolled in an accredited Indiana school during the Spring IREAD-3 test window and/or before the end of the 3rd grade school year are required to pass IREAD-3** (see Question 1 of this FAQ for more information).
28. How and when should schools / corporations notify IDOE of any students whose IEPs or ILPs require a paper form of the assessment?	If a school is testing online, a paper form of the assessment can be provided only if it is formally documented as an accommodation in the student's educational record. The paper form of the assessment can be ordered by the school corporation in the DOE-Test Layout (TL) prior to the testing window.
30. How can I access legislative and State Board action regarding IREAD-3?	House Enrolled Act (HEA) 1367, also known as PL 109 in 2010: http://www.in.gov/legislative/bills/2010/PDF/HE/HE1367.1.pdf Citations for the statutory language created in HEA 1367 (2010) are in IC 20-32-8.5: http://www.in.gov/legislative/ic/2010/title20/ar32/ch8.5.html The State Board Rule in the IAC is found under Rule 3.1: http://www.in.gov/legislative/iac/T05110/A00062.PDF The final rule as published in the Indiana Register: http://www.in.gov/legislative/iac/20110420-IR-511100635FRA.xml.pdf

Source: (IDOE, n.d.)

Appendix D

Additional Tables

As a point of comparison, Tables 18 and 19 show the same subgroup comparison as Tables 8 and 9 but after the dataset was "cleaned" by removing any records with incomplete data including students who did not have reported scores through the entire run from Grade 3 to Grade 8. The results show some differences in the descriptive statistics once the incomplete/invalid records were removed, but there are not any glaring results.

Table 20 shows the effects of retention on ELA and Math scores using the same lmer function in R to do longitudinal multilevel modeling when adjusted to run on a dataset where all students with IEP or ELL exemptions are removed. Therefore, the model was also slightly altered by removing IEP and ELL as factors.

Table 18

Same-Grade Comparison of ISTEP+ English / Language Arts Standardized Test Results (as z-scores) for 2012 Grade 3 Cohort After Cleaning the Dataset

Grade	Academic Year		Did Not Pass IREAD-3 and Promoted	Did Not Pass IREAD-3 and Retained	Passed IREAD-3 and Retained	Passed IREAD-3 and Promoted
Grade 3	2012	N	3388	931	322	58662
		Mean	-1.142	-1.614	-1.132	.122
		Median	-1.318	-1.557	-1.097	.064
		S. D.	1.280	.870	.662	.879
Grade 3 (post-retention)	2013	N	---	933	326	---
		Mean	---	-1.246	-.992	---
		Median	---	-1.285	-1.046	---
		S. D.	---	.687	.488	---
Grade 4	2013 or 2014	N	3381	889	319	58319
		Mean	-1.193	-.976	-.766	-.237
		Median	-1.226	-1.001	-.782	-.314
		S. D.	1.033	.660	.520	.871
Grade 5	2014 or 2015	N	3356	934	324	58165
		Mean	-.839	-.492	-.222	.012
		Median	-.860	-.416	-.217	-.017
		S. D.	.846	.624	.334	.692
Grade 6	2015 or 2016	N	3451	935	322	58220
		Mean	-.671	-.631	-.268	.306
		Median	-.636	-.631	-.276	.342
		S. D.	.604	.608	.577	.631
Grade 7	2016 or 2017	N	3396	935	323	57915
		Mean	-.866	-.509	-.188	.609
		Median	-.823	-.495	-.170	.612
		S. D.	.712	.619	.653	.805
Grade 8	2017 or 2018	N	3392	940	324	57782
		Mean	-.796	-.162	.104	.832
		Median	-.791	-.128	.136	.851
		S. D.	.822	.739	.657	.922

Table 19

Same-Grade Comparison of ISTEP+ Mathematics Standardized Test Results (as z-scores) for 2012 Grade 3 Cohort After Cleaning the Dataset

Grade	Academic Year		Did Not Pass IREAD-3 and Promoted	Did Not Pass IREAD-3 and Retained	Passed IREAD-3 and Retained	Passed IREAD-3 and Promoted
Grade 3	2012	N	3488	949	326	58474
		Mean	-.851	-1.373	-1.133	.126
		Median	-.945	-1.416	-1.126	.115
		S. D.	1.135	.837	.760	.923
Grade 3 (post-retention)	2013	N	---	931	325	---
		Mean	---	-1.220	-1.021	---
		Median	---	-1.292	-1.051	---
		S. D.	---	.810	.670	---
Grade 4	2013 or 2014	N	3419	896	322	58240
		Mean	-.986	-.866	-.765	-.243
		Median	-.936	-.892	-.803	-.313
		S. D.	.860	.677	.653	.888
Grade 5	2014 or 2015	N	3390	934	324	58205
		Mean	-.716	-.316	-.224	.130
		Median	-.651	-.284	-.224	.124
		S. D.	.807	.448	.421	.760
Grade 6	2015 or 2016	N	3451	933	323	58220
		Mean	-.442	-.346	-.227	.373
		Median	-.416	-.334	-.244	.382
		S. D.	.530	.603	.573	.505
Grade 7	2016 or 2017	N	3410	935	321	57585
		Mean	-.266	-.137	-.052	.610
		Median	-.274	-.135	-.047	.594
		S. D.	.555	.551	.565	.688
Grade 8	2017 or 2018	N	3417	940	324	57607
		Mean	-.089	.177	.330	.937
		Median	-.062	.235	.389	.946
		S. D.	.716	.612	.531	.685

Table 20

Summary of the Effects of Retention in ELA and Math Using Longitudinal Multilevel Modeling with School District Nesting (IEP and ELL exemptions removed)

	Estimate	Std. Error	t-value	95% Conf. Interval	
				Lower Bound	Upper Bound
ELA Z-scores					
Retained	0.212*	0.040	5.296	0.134	0.290
Math Z-scores					
Retained	0.180*	0.048	3.777	0.086	0.275

*Notes. * The mean difference is significant at the .05 level.*