The Dissertation Committee for Man Yang Certifies that this is the approved version of the following dissertation:

The Effect of Grade Retention on Academic and Social-behavioral Outcomes for Students with Disabilities in Elementary Grades

Committee:

North A. Cooc, Supervisor

Marcia Barnes

Soyoung Park

Jessica R. Toste

James E. Pustejovsky

The Effect of Grade Retention on Academic and Social-behavioral Outcomes for Students with Disabilities in Elementary Grades

by

Man Yang

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2018

Dedication

This dissertation is dedicated to my mom, dad and husband: I could not have completed it without their encouragement, love, patience, and unconditional support. This accomplishment is as much theirs as it is mine.

Acknowledgements

First and foremost, it is beyond argument that the most influential person in my graduate career has been my doctoral supervisor, Professor North Cooc. I thank him wholeheartedly for his contributions of time, ideas, and funding; making my Ph.D. experience memorable, productive and inspiring. His guidance and discipline have been indispensable to my growth as a researcher and as a person over the past four years.

I am also indebted to my committee members, without whom my research work would not have been possible. Professor Marcia Barnes has done a fantastic job of providing insights in the discussion of my paper and I am continuously awed by her meticulousness. Professor James Pustejovsky has been a constant and reliable source of support on methodology. He taught me everything from homoskedacity, to what hierarchical linear modeling is, to the advanced techniques of causal inferences. His rigorous academic standards have been critical. Professor Jessica Toste was kind enough to share her expertise so willingly and undertake the editing of my work. Professor Soyoung Park, the most recent addition to my committee, has been providing numerous inquiries on the theoretical significance of my work. Her remarks have meant substantially to me and the field. In all seriousness, all of my committee members have been graciously generous with their time, ideas and connections. Thank you all.

My intellectual debt also goes to those remarkable colleagues, whom I always look up to. I have had the great pleasure of working on the project led by Professor Sylvia Linan-Thompson when I was admitted to UT in 2014. Priceless memories of early morning carpools with Maria to local schools collecting data and late afternoons scoring countless piles of test booklets with Alfredo remain vividly emblazoned in my mind till today. I also managed to meet people outside of the department. I would like to express the deepest

appreciation to Li Sheng, a former professor from the College of Communication, who granted me chances to work with her on different projects. I had the great fortune of becoming close friends with two of her doctoral students Boji Lam and Ying Hao, who've scrapped me off the ground and reassembled me into a human again.

My time at UT was made delightful in large due to the many friends that became a part of my life. Thank you for your friendship, sense of humor, and endless support along the way! Shout out to Le Tran Ozor, Alexis McCoy, Laci Watkins, Kelly Williams, and Chelseaia Charran, who infuse monotonous Ph.D. life with laugh and joy. I become more and more grateful to Luping Niu as I write up my dissertation as I'd still be trying to figure out how to crunch data if it weren't for him. Most people are lucky to have one or two friends as close to their heart as Han Li and Karlie Hoh are to mine. I could not possibly have made through some of the rough times over the past few years without their constant love and support. I really appreciate Han's candor with me about everything and Karlie's company on years of coffee runs at the Mozart Coffee by Austin lake. I will say to both Han and Karlie, I love you and your support has been treasured dearly. There are no words to express my gratitude to Kathy Strunk, a dear old friend of mine at Nashville, who devotedly supported me over the Ph.D. journey. She taught me to become a person true to myself and I wish her all the best in the future.

I would like to extend my sincere thanks to my colleagues at Learning Policy Institute. It was a great honor to join the group while finishing up my dissertation. I shall never forget the day when I sweat over the entire interview with Linda. LPI has been a source of friendship as well as good collaboration. To thank each of my colleague at LPI would be overwhelming, but they are all truly appreciated. Special mention to Dion, who set aside time albeit millions of works on his plate and helped congregate colleagues as

audiences for my defense rehearsal. I also owe my deepest gratitude to Patrick, my supervisor at LPI, who generously offered support throughout the final stage.

Finally, but by no means least, I would like to thank my family for all their love, support and encouragement. For my parents who raised and supported me in all my pursuits. And most of all for my encouraging husband Haoran whose faithful support, wise counsel and sympathetic ear during all stages of this Ph.D. is deeply appreciated. Thank you.

Abstract

The Effect of Grade Retention on Academic and Social-behavioral Outcomes for Students with Disabilities in Elementary Grades

Man Yang, Ph.D.

The University of Texas at Austin, 2018

Supervisor: North A. Cooc

Abstract: There is a lack of research examining the rates and effect of grade retention, a widely used intervention at schools, among students with disabilities. The existing research evidence of retention effects among students without disabilities remains inconclusive, which might result from the differences in analytical methodologies, sample selections and measurement errors. The purpose of the present study was to examine the retention rates, academic and behavioral trajectories of retained versus promoted students with disabilities, and the effect of grade retention for students with disabilities using the nationally representative Special Education Elementary Longitudinal Study (SEELS) dataset. Hierarchical linear modeling (HLM) was conducted to investigate the trajectories of being retained at three different time points (kindergarten, first and second grade) on later academic and social development trajectories among students with disabilities (N = 13,176). Propensity score matching (PSM) was used to infer causal inferences by creating two equivalent groups matched on a set of baseline measures. Results indicated that retention rates for students with disabilities are much higher than that for students without

disabilities reported in previous studies. In addition, retention rates differ, albeit slightly, by student demographic characteristics such as race, gender, EL status, family income and disability types. Retained students with disabilities do not differ significantly from their promoted peers in terms of academic and behavioral trajectories with an exception of mathematics at kindergarten. The analysis of propensity score matching indicated that there is no effect of grade retention on reading performances but retention results in significant worse math performances for students with disabilities.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

The National Center for Education Statistics (2009) estimated that, between 1996 and 2007, between 9% to 11% of students in kindergarten through grade 8 had been retained at least once. Grade retention is widely used in schools for students who fail to meet grade-level expectations (Range, Dougan, & Pijanowski, 2011). Other reasons for grade retention include a lack of social maturity, failure to meet criteria for promotion, and frequent unexcused absences (Bowman, 2005). A question that has been frequently asked by researchers is whether retaining students who failed to meet grade level standard is correct and wise, and does retention benefit all students in terms of their academic and behavioral development? Considering high stakes testing and greater pressure put on school districts to ensure all students, including students with disabilities, are making enough progress to meet grade level standards, grade retention has become a controversial topic faced by schools and teachers (Larsen & Akmal, 2007; Scott, 2011). The school accountability system and the desire to raise academic standards has made grade retention a seemingly remedial strategy for school practice and teachers often times choose to retain children without a solid understanding of its effects, especially when retention was used among students with disabilities (Renaud, 2013). Grade retention is also a costly intervention, for example, in the 2010 to 2011 school year, Texas state spent approximately $7.2 billion for an extra year of schooling for a total of 826,367 children retained in grades kindergarten through twelfth grade which accounts for 16.6% of all school aged students in Texas (Texas Education Agency, 2012).

Studies have found the number of students retained varies by states, years and grade levels (Hauser, Frederick, & Andrew, 2006; Warren, Hoffman, & Andrew, 2014; Warren & Saliba, 2012). There are also substantial differences in rates of grade retention by student subgroups defined by gender, race, parental education and family socioeconomic status and rates are higher among boys, minorities, those who have mother with lower education degrees and those who are from poor families (Warren et al., 2014).

Despite the wide heterogeneity in grade retention among students, little is known about the retention rates of students with disabilities, a population perhaps most at risk of repeating grades. Currently nearly 13 % of the school-aged population receives special education services under the Individuals with Disabilities Education Act (2004). Studies on grade retention indicate that lower academic achievement and greater behavior problems are strong predictors of grade retention among students without disabilities (Willson & Hughes, 2006; Winsler et al., 2012). The majority of students with disabilities perform well below grade levels, showing little improvement in national assessment on reading and math achievement. For instance, the National Assessment of Educational Progress (NAEP) shows that only 32% of fourth graders with disabilities score at or above a basic level on reading, compared with 73% of students without disabilities; and 31% of eighth graders with disabilities scored at or above basic levels on math, compared with 75% of students without disabilities (NAEP, 2017). Therefore, it is within our reasonable belief that students with disabilities may be at higher risk of being retained

although very little information has been given to report retention rates on this particular student population. Studies also suggest that grade retention may benefit students who are excessively absent during the school year. Thus, for students with disabilities, who are more likely to miss school, retention may have a positive effect (Jimerson & Renshaw, 2012). Taken together, given that students with disabilities are more vulnerable to educational inequality and retention has been used as a remedial strategy by schools especially for students who are not meeting grade level expectations, it is imperative to get a better idea of how many students with disabilities were retained and whether retention help improve their educational outcomes.

Information about how many within this population are retained each year and whether there is an impact of retention on their educational outcomes is scant. Lorence and Dworkin (2006) investigated grade retention in Texas using data from the Texas Education Agency (TEA) and reported that the retention of students with disabilities is higher than that for students without disabilities, except for Black students. A more recent study by Tingle, Schoeneberger and Algozzine (2012) investigated a southeastern regional dataset from two consecutive academic years and they concluded that the majority of the retained students were boys, minorities, students with disabilities, and English language learners. About 5.4% of students with disabilities were retained, whereas the retention rate for students without disabilities was 2.3%. These are the only two studies that have directly examined the distribution of grade retention by special education status. However, both studies have limitations in that they used regional data

and reported retention rates among students with disabilities as a whole without further disaggregating on along other student characteristics for this population. Therefore, we know little about the retention rates among students with disabilities on a national level and whether the retention rates in this particular group of students differ by student demographic characteristics. Such information could better inform retention policies and interventions to improve long-term outcomes for students with disabilities.

There is also little information on grade retention effects on academic and behavioral outcomes for students with disabilities. Previous research examining grade retention effect on student academic outcomes among general education populations have produced mixed effects for retained students when compared to promoted peers (e.g., Chen, Hughes, & Kwok, 2014; Hong & Raudenbush, 2005; Shane R Jimerson, 1999; Lorence & Dworkin, 2006; Moser, West, & Hughes, 2012; Walters & Borgers, 1995). It is also unclear whether grade retention may be detrimental to the social and emotional development of students with disabilities, a population that is the focus of this study (e.g., Tingle, Schoeneberger, & Algozzine, 2012; Yamamoto & Byrnes, 1987). Some previous research show students make gains by the end of the first year after being retained and these gains may continue to the second year (Peterson, DeGracie, & Ayabe, 1987; Walters & Borges, 1995). Retention may also improve peer relationships in that an extra year may provide more opportunities to be familiar with routine activities and make friends (Gleason, Kwok, & Hughes, 2007; Wu, West, & Hughes, 2010). Critics of retention argue that retained children are deprived of opportunities to engage in age-

appropriate cognitive and social activities that can hinder the development of academic interests, self-regulation, and interpersonal skills (Morrison, Griffith, & Alberts, 1997). For instance, researchers have found that retention negatively affects self-esteem because students often times view retention as a penalty (Brophy, 1983; Cochan & Qadir, 2004). Students who are retained are also less likely to graduate from high school than their promoted peers (Stearns, Moller, Blau, & Potochnick, 2007). In the long-term, there is evidence that students with a history of retention have higher chances of being unemployed or committing a crime (Venable, 2015). Whether these trends are consistent for students with disabilities have not been examined.

In addition to debates about whether retention is beneficial, previous studies have also found mixed evidence regarding when retention should occur, or whether the timing of retention makes a difference. For instance, some researchers argue that the benefit of grade retention is most valuable in cases of child immaturity and excessive absences. In these cases, children with slower developmental trajectories are in need of more time to develop skills necessary to proceed to the next grade level (Dougan & Pjanowaski, 2011; Jimerson & Renshaw, 2012; Hong & Yu, 2008). Retention is also found to be more effective for young children in early primary grades to prevent learning challenges from becoming more severe later on (L. A. Shepard & Smith, 1989). An exception is academic redshirting, where students are purposely delayed kindergarten entry to prepare for school, there is less evidence of benefits for young children (Graue & DiPerna, 2000;

Martin, 2009). However, for students with disabilities who have greater academic and social challenges, early grade retention may be effective.

Taken together, previous research on grade retention suggests that the retention trend varies across different student populations, and the impact of retention on academic and social outcomes is mixed. More importantly, there is a large gap in the literature on students with disabilities. Research that examines grade retention rates and its effect on students with disabilities has the potential to inform school policies aimed at raising achievement for all students. In addition, this result can inform budget decisions given that retention is a costly intervention for schools (Bowman, 2005). Therefore, a more thorough understanding of grade retention effects on students with disabilities is critical to justify the delayed educational opportunities associated with retaining students with disabilities to produce better academic and social outcomes.

This study is one of the first to examine grade retention among students with disabilities. Using nationally representative data from the Special Education Elementary Longitudinal Study (SEELS), I addressed the following three research questions: (1) What is the grade retention rate for students with disabilities overall and by student demographic characteristics in elementary school from kindergarten to fifth grade? (2) What are the academic and social-behavioral trajectories of students retained from kindergarten to second grade? (3) Does the grade of retention make a difference on academic achievement for students with disabilities?

CHAPTER 2: REVIEW OF THE LITERATURE

GRADE RETENTION THEORIES

Grade retention is defined as repeating a grade or spending an extra year of schooling completing the same grade level (Bowman, 2005). The primary reason for retaining students is the failure to grow academically or socially (Venable, 2015). A lack of maturation or the failure to meet grade-level expectations are the two driving forces behind a student needing to repeat a grade. There are several opposing theories that were used by prior researchers to support or against the use of retention when maturity is the primary focus. Scholars draw from social comparison theory (Festinger, 1954) to advocate for the use of grade retention. Social comparison theory states that "people tend to move into groups which, in their judgment, hold opinions that agree with their own and whose abilities are near their own" (Festinger, 1954, p.136). Placing students in a context where his classmates and himself are at similar or the same level of maturity creates more opportunities to learn and thrive. This theory is in accordance with the nativist view in that maturity is the basis to determine whether a child is ready for school. Therefore, retention provides students with more exposure to academic content and to become more knowledgeable and competent in peer relationships (Plummer & Graziano, 1987). In line with this thinking, teachers and parents who support the use of retention believe that young children who fail to reach the academic or behavioral norms of schools need more time to mature before moving to the next grade (Smith & Shepard,

1988). They argue that the consequence of promoting children before they are ready is they may fail academically and socially in the long term. On the other hand, labeling theory opposes the idea of grade retention because of the potentially detrimental effect on children's development, particularly on self-esteem and self-perceived competence (Becker, 1963; Lemert, 1967). The negative label associated with grade retention might result in a rejection of the retained children by their peers or teachers (Pagani, Tremblay, Vitaro, Boulerice, & McDuff, 2001). Moreover, opponents of retention also argue that the idea of repeating a grade ignores the interactive effects of environmental influences on human development (Morrison, Griffith, & Alberts, 1997). They argue that placing children in age-inappropriate settings might prevent them from growing cognitively and socially, while simultaneously depriving them of opportunities to engage with peers their age. Similarly, the interactionists believe children's readiness relies not only on their learning skills but also depends on the capacity of schools to meet children's needs (Meisels, 2002). Considering the controversial and critical role that grade retention plays in student learning, it is vitally important to understand how many students are retained annually, who those retained students are and how retained students perform.

Given grade retention is underpinned by those theories related to child development and how a child is deemed to fit with his or her cohort seems to be the focal point of the argument, both social comparison theory, nativist view, labeling theory and interactionists view are potentially relevant to decisions of retaining or promoting students with disabilities (Martin, 2009). The labeling theory can be seen as a basis for

grade promotion because students with disabilities were pushing into the "normal" range addressed by the school without being placed on labels that may result in negative consequences to their development. From the perspectives of interactionists, grade retention should also be avoided because "there is a great range of normal variation in development that can be accommodated by the school" (Martin, 2009, p.4). On the contrary, social comparison theory and nativist view can be seen as the basis to retain a child with disabilities if they do not fit with the environment resulting from immaturity or school failure. For students with disabilities, the driving force of the decision between grade retention and promotion lies within how grade retention affect student outcomes when the targeted student population generally performs at a lower academic level and encounter more social-behavioral challenges. The present study is the first to shed light upon which of the two theories is more pertinent to grade retention and its effect on academic and social-behavioral development for students with disabilities.

TRENDS IN GRADE RETENTION RATES AND MEASUREMENT

Grade retention is widely used in public schools as a remedial intervention in response to states' accountability systems. The relationship between grade retention and academic achievement, social skills, and high school completion has been the subject of numerous empirical studies over the last couple of decades (e.g., Chen, Hughes, & Kwok, 2014; Jimerson, 1999; Hong & Yu, 2008; Moser et al., 2012; Walters & Borgers, 1995; Wu, West, & Hughes, 2008). The majority of the existing studies tend to report retention

rates among all students as a whole and report disaggregated rates by student demographic characteristics such as gender, race and family income. Nevertheless, there is surprisingly little information about how many students with disabilities repeated grades overall and what the demographics of those retained students with disabilities are across the nation. Further, there is no current systematic way to quantify grade retention rates. Existing measures of retention rates on state and national level suffer from the use of outdated data, a lack of direct measurement about individual retention status, and data sampling biases. Studies that use indirect measurement of retention such as proxies indicate retention rates differ by state, academic year, grade level, gender, student race, and family income levels. For example, retention rates are higher among early primary (K-2) and high school grades than other grades. At the national level, retention rates have declined from 1994 (2.7%) to 2010 (1.5%) using the 1995-2010 Current Population Survey (CPS) data (Warren, Hoffman, & Andrew, 2014). If disaggregated by state, states such as Texas and Alabama had higher retention rates than the national rate, whereas other states such as Oregon and Arizona had lower retention rates (Warren & Saliba, 2012).

The most recent national data from the National Center for Education Statistics (NCES) on annual grade retention rate dates back to 2009. According to this report, the retention rates for students from kindergarten through grade 8 has remained between 9 to 11% between 1996 and 2007. The report shows retention rates differ by student characteristics, such as gender, race, family socioeconomic status, and maternal education level. For instance, about 12% of male students had been retained for at least one year,

compared with 8% of female students; 23% of students from poor families had been retained at least once, compared with 11% of students from near-poor families and 5% of students from non-poor families; 20% of students whose mothers had less than a high school diploma had been retained for at least one year, compared with 3% of students whose mothers' highest level of education was a bachelor's degree or graduate school; and, 16% of Black students and 11% of Hispanic students had been retained at least once, compared with 8% of their White peers (NCES, 2009). But nowhere in the report are retention rates among students with disabilities mentioned, another student characteristic that marks a particular population who are more likely to be retained considering their lower achievement and who are more vulnerable to education equity.

The 2013 to 2014 Civil Rights Data Collection (CRDC) is by far the most recent national database that captures retention information from all of the nation's public schools. The report shed some light on retention rates of students with disabilities. According to the national estimations, about 19% of retained kindergarteners are students with disabilities; the percentage goes down about three to four percentage points as grade level increases. By fifth grade, about 17% of retained students are students with disabilities. Retention rates were also reported to vary by student race. For instance, about 46% of retained kindergarteners are white, followed by 25% Hispanic and 21% Black. By second grade, about 35% of retained students are Hispanic, followed by 31% Black and 29% white. More Black students were retained starting from third to fifth grade. However, given the nature of the dataset, it does not separately report retention rates among students with and without

disabilities, nor does it further investigate retention rates among students with disabilities by student background characteristics. Hence, there's little information available to compare whether students with disabilities are at higher risk of being retained as opposed to their typically developing peers.

Previous researchers have used different measures to quantify retention rates, such as using proxies for retention based on the distribution of students' grades of enrollment conditional to their ages (e.g., Bianchi, 1984; Frederick & Hauser, 2008; Heubert & Hauser, 1999). The proxies define students as retained if they are enrolled below the modal grade for their age. For instance, if 7% of 8-year-olds were enrolled below their modal grade (third grade) in the fall of a given year, and 8% of 7-year-olds were enrolled below their modal grade (second grade) in the fall of the previous year, then one could infer a 1% second-grade retention rate at the end of the given year. Using this proxy, Hauser and colleagues (2006) examined grade retention rates with the October school enrollment supplement to the Current Population Survey (CPS) from 1996 to 2003. The CPS school enrollment data come from a nationally representative sample of the civilian non-institutional population each October over an extended period of time.

Similar to the NCES (2009) report, Hauser and colleagues (2006) also found retention rates differ by grade levels, gender, race and family income levels. About 4.5% of kindergarteners were retained and in first grade, about 7% of students were retained. The rates decreased by about 67% in second grade when 2.3% of students were retained. The rates then hover between one and two percent until eighth grade with an exception of

seventh grade when the rate was about 2.3%. Ninth graders had the highest retention rates at 3.1%, which is still less than half of the retention rate for first graders. In tenth and eleventh grades, the rates dropped back to around 2%. Retention rates were also found to be fairly consistent across time. For instance, the retention rates between 1996 to 2003 were between 2.3% to 3%. Males had greater retention rates than females; about 2.8% of boys were retained in 1997, whereas less than 2% of girls were retained in the same year. Blacks had the highest retention rates (3.8%) in 1996, compared to Hispanics (3.3%) and of Whites (1.8%) and other races. Students from the lowest income families had the highest retention rates across years. About 10% of kindergarteners from the lowest income families were retained, compared with 4% of their peers from the highest income level families. The gap persisted until eleventh grade when 4% of students from the lowest income families were retained as opposed to 1% of their peers from the highest income level families. However, the retention rates by grade level, gender and family income reported by Hauser and colleagues are much lower than that reported by the NCES (2009). Possible reasons to explain the discrepancies include the use of different dataset, different span of years, and different methods to calculate retention rates. For instance, NCES report used the National Household Education Survey Program (NHES) from 1996 to 2007 and retention information was collected from parent questionnaires.

Another method used to calculate retention rate was described by Warren and Saliba (2012). In their study, they used information about enrollments and population size to calculate national and state-level retention rates from academic years 2002 to 2009

using NCES's Common Core of Data (CCD). For example, the retention rate for first grade can be determined using three pieces of information: first grade enrollments in two consecutive years and the number of first-time first graders in the second year. Accordingly, the retention rate for second grade can be calculated using second grade enrollment, first grade retention rate, and first grade enrollment. If there are 1,111 first graders enrolled in Fall of year X and among them 1,000 are first-time first graders, the retention rate for first grade could be produced using [(1,111-1,000)/1,111] * 100 = 10%. Using the first-grade retention rate, if there are 1,020 students enrolled in the fall of year X, the second-grade retention rate could be retrieved by [1,020 – (1-10%)*1,111]/1,020 * 100 = 2%. Using this approach, the authors calculated the retention rates for first through eighth grade from 2002-2003 and through 2008-2009 for the entire nation. They found retention rates were the highest in first grade (3.5%) and that retention rates decline slightly across grade levels. They also compared retention rate by grade level for all states. For example, at the end of the 2008-2009 academic year, the national retention rate was about 3.5%, lower than that in Tennessee (3.8%), New York (4.6%), Florida (5%) and Texas (4.8%). In contrast, states like Oregon (1.5%), Washington (1.8%) and Michigan (3.3%) had lower than national average retention rates. Apart from national and state-level retention rates, Warren and colleagues (2014) also examined the retention rates using the method disaggregated by grade level, gender, race and family income levels using CPS data from 1995 to 2010. Their results regarding the percentages associated with the aforementioned student characteristics are in consistent with what

Hauser and colleagues (2005) found in their study that retention rates are higher among early primary grades such as kindergarten through second grade, males are having higher rates than that of females, minority students were retained more than their white peers, and more students from poor families were retained. With the discussion on the different ways to quantify retention rates, it is clear that in order to get more accurate estimate on how many students are retained as well as by student demographic characteristics, more recent data collected at national level with individual level information on retention is necessary.

Regardless of the differences in methods used to calculate retention rates, what is still overlooked in previous studies are students with disabilities. In one of the few studies to include this population, McLeskey and Grizzle (1992) investigated the grade retention rate in Indiana during the 1987-1988 school year using the data for students with learning disabilities in the state. They found 58% of the students with learning disabilities were retained, which is almost twice as many as students without disabilities. Among students with learning disabilities, those who were retained were referred for special education one year later than their promoted peers. In addition, in deciding whether to retain a student with a learning disability, they found the absolute achievement level of the student largely outweighs the difference between his or her expected and actual achievement level. Tingle, Schoeneberger and Algozzine (2012) also analyzed data on students with disabilities from a large school district located in the southeastern region

and reported about 5.4% of students with disabilities were retained compared with an overall retention rate of about 2.3% for students without disabilities.

These two studies serve as the only source of information on the grade retention rate for students with disabilities and both have several limitations. The data used in the first study is from only one state and focused exclusively on students with learning disabilities. In addition, the first study (McLeskey and Grizzle, 1992) was conducted before major educational reform such as NCLB and IDEA were implemented, which largely compromises relevance of its results. Similarly, the second study only used regional data and may not be nationally representative. Taken together, due to the large variations existing in the use of data sources, methods used to calculate retention rates, focus on different student populations, and time of the publications, the retention rates reported by the previous studies and reports are unsurprisingly varying in a wide range. More attention should be directed to the exploration of grade retention rates among students with disabilities using nationally representative data and including and disaggregating for more disability types. What also remains missing from both studies is the investigation of the impact of retention on students with disabilities in terms of their academic and social outcomes, which is addressed in the current study.

GRADE RETENTION AND STUDENTS WITH DISABILITIES

The concern with grade retention is how it may affect student academic and behavioral achievement. Studies on the general education population have been well

documented but the results are inconsistent. There is also little information on how grade retention affects students with disabilities, a student population perhaps most at risk for repeating grades given students with disabilities are more likely to struggle academically and socially. Mantzicopoulos (1997) investigated 40 children with attention problems to examine whether pre-elementary retention results in positive long-term academic outcomes. The results fail to support the use of early retention for children with academic or behavioral difficulties as retained children's ratings on reading and behavioral dimensions declined, whereas their promoted peers' ratings increased over time. However, their findings found no consistent differences between the two groups in terms of math achievement, indicating retention effects may differ by subject areas. Additionally, the study had a small sample size and there was no statistical control and adjustment of the two comparison groups regarding student background or prior achievement, which may result in an inaccurate estimate of the retention effect. This study is the only one that has linked grade retention to outcomes for students with disabilities.

Although there are very few studies examining the effects of grade retention on students with disabilities, the existing literature on students without disabilities point to potential hypotheses. Previous studies show that grade retention may benefit students who are excessively absent during the school year. For students with disabilities who are more likely to miss schools because of their disabilities, retention may be a promising strategy for them to catch up academically. In addition to excessive absence, retention was also found to most likely to benefit students who need more support in their academic career or

who are lacking necessary learning related skills such as attention, behavioral and emotional self-regulation, and social skills (e.g., Chen et al., 2014; Hong & Raudenbush, 2005; Renaud, 2013). Students with disabilities could fall within this group considering the challenges they are facing at school that come from both their disabilities and the instruction they receive.

For the remainder of the chapter, studies of grade retention effect on student academic and behavioral outcomes among the general education population are reviewed. Because of the differences in data source (i.e., national or district data, school or student level data), sample of participants (i.e., grade level, particular student population), research design (i.e., longitudinal or cross-sectional, observational or experimental), analytical methods (i.e., PSM or other) and types of outcomes (i.e., short- versus long-term, academic versus behavioral), it is not surprising that study findings regarding the retention effect differ. Nevertheless, the review of existing studies that focus on grade retention among students without disabilities could help inform the current study of the appropriate study design and reliable analytical methodology in order to produce an accurate estimate of the grade retention effect on academic and social or emotional outcomes for students with disabilities. The following sections start with a discussion of methodological issues in some of the previous studies. Next, I review studies that examined impacts of retention on academic outcomes followed by social or emotional findings and within academic and social outcomes.

Methodological Issues

Methodological and design quality of studies are believed to have direct impacts on observed effects (Allen, Chen, Willson, & Hughes, 2009). A major limitation with some of the earlier studies of grade retention is an inability to make causal inferences about effects of retention because these studies did not employ a randomized experimental design (Campbell & Stanley, 1966). However, in the case of grade retention it is neither feasible nor ethical to conduct a randomized trial. In the absence of random assignments (i.e., to a retention versus promotion condition), researchers mainly reply on two types of methods to control for pretreatment differences: one is the equating method to select a comparable group of low-achieving but promoted students to compared to retained students and the other method is to use statistical controls. The first method primarily relies on the fact that the two groups do not differ significantly on pre-assignment measure(s). However, this design does not guarantee the two groups are fully equivalent on the measured variables and it is also possible that the two groups are not equal on other unassessed attributes that are related to outcomes (Heinsman & Shadish, 1996). To compensate for the limitations in the first method, statistical control helps account for the potential selection bias and can produce better estimate of the effect of retention. As Moser and colleagues (2012) noted, "studies that do a better job of controlling from student characteristics associated with selection into the retention are less likely to find that grade retention has a negative effect

on achievement" (p.603). Failing to account for non-random differences between retained and promoted peers will likely lead to biased estimates of the retention effect on student outcomes. Hence, in order to accurately examine the effects of grade retention on student performance, statistical models need to take into account observed and unobserved factors associated with student performance and whether students are retained or promoted.

One statistical control method that has been widely used is propensity score matching (PSM). Propensity score analysis offers important benefits over the previously used equating method in that it substantially reduces selection bias if properly used (Rosenbaum, 2002; Rosenbaum & Rubin, 1983). Propensity scores are the estimated probabilities of being assigned to the treatment group, in other words the retained group and PSM is a technique used to equate the promoted and retained groups on the basis of pretreatment variables.

Retention Effect on Academic Outcomes

Scholars have yet to reach an agreement on the findings that relate to the effect of retention on student academic outcomes given the mixed evidences presented in prior studies. This may be a result of whether a rigorous study design was applied by using statistical controls to account for pretreatment differences among participants. Positive evidences of grade retention effect have been identified in more recent methodologically rigorous studies. In a study where the causal effect of grade retention on high school completion in Chicago was assessed using regression discontinuity (RDD), the authors

concluded retaining low-achieving sixth or old eighth grade students had no impact on their high school completion but retention substantially increases the probability of dropping out among young eighth graders (Jacob & Lefgren, 2009). This finding further suggested the effect of grade retention effect on later outcomes may depend on the grade and achievement level of students (Dawson, 1998; Isenhart & Bechard, 1987). Positive retention effect was also identified in studies that use PSM. Propensity score-based adjustment methods identify control units that can be matched with the treated units on the estimated propensity scores so that the unbiased estimates of the treatment effects can be produced. Researchers believe it is possible for PSM to remove baseline differences between the treatment and control groups if important variables related to both treatment selection and outcome were identified. In such case, balance on baseline levels of each measure variables could be achieved with the proper use of PSM. Wu and colleagues (2008) found that retained children at first grade produced different growth rates in reading and math. In addition, they also found a slower increase in academic achievement in a short-term period defined as the first two Waves which covers from approximately 0.5 year before retention to 0.5 year after retention. However, the retained children produced a faster growth rate of reading than their promoted peers in the long-term period defined as the latter three Waves which covers from roughly 0.5 year after retention to 2.5 years after retention. More interestingly, the study also identified moderators of the retention effect, such as home-school relationship and children's externalizing problems. Children with lower levels of externalizing behaviors benefited more from promotion than retention in the short term,

whereas children without strong parent involvement in their education had greater short-term academic benefit resulting from retention versus promotion. Short-term positive effects were also documented in Moser and colleagues' study (2012) in which they investigated retention effects on first graders by examining their growth trajectories in math and reading performances up until fifth grade. The authors identified a 1-year boost in achievement among retained children but this initial advantage was fully dissipated by the end of elementary school. Moreover, they also concluded that early grade retention can protect children from being retained later on but it does not decrease the probability of retained children being placed in special education. In fact, both retained and promoted children did not differ in their probabilities of being referred for special education.

Besides recent rigorous studies in which positive retention effects were found, earlier studies that did not statistically control for possible pretreatment differences also found positive retention effect. For instance, Rust and Wallace (1993) found students made significant academic gains over the year they were retained, although the gains plateaued during later years. Jimerson and colleagues (1997) concluded children retained once in kindergarten through third grade experienced academic gains in math compared to the control group consisted of randomly selected subjects who were promoted. It was also during the retention year that the retained students made significantly higher academic gains than their promoted counterparts that continued through the second year (Peterson, DeGracie, & Ayabe, 1987). Kerzner (1982) also documented the benefits of grade retention among students from grades 1-5 who had been retained once and found students' test scores

increased significantly during the retention year. She further suggested that retention may be beneficial for students at lower grades and that an advantage could possibly be sustained till fifth grade. It should be noted that the studies that tend to find positive retention effects primarily focus more on the short-term or temporal academic outcomes among students in the early grades.

Conversely, negative effects of retention have also been documented in rigorous studies using PSM. Hong and Raudenbush (2005) analyzed data from the *Early Childhood Longitudinal Study Kindergarten* cohort (ECLS-K) using multilevel propensity score stratification and found significant average negative effect of repeating kindergarten on children's cognitive growth in reading and math. In addition, retained children were found to lose an average of two-thirds of a standard deviation in each subject area, indicating they would have learned more had they been promoted. The authors concluded that kindergarten retention in general might prevent children from learning adequately, except for those who were at risk for academic failure and most likely need to be retained because very few promoted children were matched with those highest-risk children. Similar findings were also presented in another study (Hong & Raudenbush, 2006) using the same dataset in which children who were retained learned less on average than did similar children who were promoted. Earlier studies with statistical control over pretreatment differences also identified negative effects associated with retention. For instance, Reynold (1992) investigated the effect of retention using data from children in Chicago tracked by a longitudinal study and he found retention had substantially negative effects on achievement

in reading and math. The finding is in contrast with what Jacobs and Lefgren (2004) found in their study although they used the same dataset. The reason may lie within the policy context of the two studies. Reynold's study was conducted prior to the implementation of standards-based accountability practices in Chicago Public Schools in the 1996 to 1997 year. According to Jacob and Lefgren (2005), school districts provided additional resources to meet the needs of retained students so that they could catch up with their peers after accountability practices were implemented. Therefore, the difference of policy contexts as well as who were may have contributed to the difference in findings of retention effects in the two studies (Allen et al., 2009).

Although less convincing, studies with less rigorous statistical design also identified negative retention effect on student academic outcomes. Using a group of low-achieving students who narrowly met the promotional cutoff, Roderick and Nagaoka (2005) examined the progress of students for 2 years after they were retained in Chicago. They concluded that students who were retained continued to struggle and faced increased probability of being placed in special education. However, retention produced different effect on student achievement at different grade level. Specifically, no evidence was found to associate retention with greater achievement growth among third graders whereas retention was found to be associated with lower achievement growth among sixth graders. Jimerson (2001) conducted a meta-analysis on retention effects and he found an effect size of -.39 for retained students based on a total of 169 effect sizes from 18 studies. Retained students had lower effect sizes in language arts (-.36), reading (-.54), math (-.49),

composite academic scores (-.20) and grade point average (-.18) than promoted students. However, among the 18 studies reviewed in his meta-analysis, only 4 used high-quality statistical controls whereas the remainder used less rigorous equating method. Therefore, the calculated effect sizes might be flawed with inaccurate estimates of retention effects in less rigorous studies.

Longitudinal retention effect was examined by student subgroups, such as those who are poorly prepared and have academic and social challenges (Chen, Hughes & Kwok, 2014). Using a sample of 530 English Learners, Hughes, Kwok, and Im (2013) found the lower expectations among parents of retained children resulted in lower student achievement and self-efficacy, indicating a negative indirect effect of grade retention on student academic outcomes although no direct relationship between retention and academic outcomes was established. Conversely, Chen and colleagues (2014) examined a group of culturally and linguistically diverse students who were retained in first grade in Texas to investigate their achievement growth trajectories from first to fifth grade. They found students with the poorest learning-related skills and academic achievement in first grade might benefit more than others from grade retention. Their findings highlight the importance of examining student subgroups when it comes to retention decisions and assessing its effect. The investigation on particular at-risk populations such as students with disabilities may compensate for what Hong and Raudenbush (2005) missed in their study. The above studies underline the importance of further examination of longitudinal retention effects by exploring other student subgroups such as students with disabilities.

Retention Effect on Social and Behavioral Outcomes

The impact of grade retention also extends to students' social and behavioral development. However, findings regarding retention effect on student social outcomes are also inconsistent, which might result from sample selection differences. Negative effects of retention on student social adjustment was documented in a few earlier less rigorous studies (e.g., Cuddy, Frame, & DeVincentis, 1987; Nikalson, 1984; Walker & Madhere, 1987). Besides social adjustment, self-esteem is another focal point of studies when investigating the social-behavioral impact of grade retention among students. Cochan and Qadir (2004) identified a negative effect of retention on student self-esteem, which could be potentially explained by students' view of retention as punishment (Brophy, 2006). In a similar vein, another study found being over-age due to retention is significantly correlated with more behavioral problems (Byrd, Weitzman, & Auinger, 1997). Retained students are more likely to skip schools, be estranged from classmates and fail to graduate from high school (Stearns, Moller, Blau, & Potochnick, 2007).

Researcher have also found long-term costs of retention on students, as well as society (Bowman, 2005). Jimerson (1999) found retained students were less likely to obtain a high school diploma, more likely to be unemployed or receive low wages, and receive less favorable job evaluations. In the meta-analysis that investigated the effects of grade retention on social and emotional outcomes, Jimerson (2001) found retained groups scored .22 standard deviation units lower than the comparison groups on a total of 77 computed

effect sizes for measures of socioemotional and behavioral adjustment. Additionally, the retained group was also .65 standard deviation units lower than the promoted group on effect sizes measuring attendance.

Recent studies applying more rigorous design tend to show a positive impact of grade retention on student social and behavioral outcome (Gleason, Kwok, & Hughes, 2007; Hong & Yu, 2008; Wu, West, & Hughes, 2010). Using path analysis, Gleason, Kwok, and Hughes (2007) found in their study that at-risk children who were retained in first grade received significantly higher peer-acceptance the following year than did their promoted peers, and the direct effect of retention on peer-acceptance was fully mediated by teacher- and peer-perceived academic competence. A more recent study by Wu and colleagues (2010) using PSM investigated retention in first grade on children's externalizing, internalizing behaviors, social acceptance, behavioral, cognitive and affective engagement revealed that retained children benefitted from repeating a grade in that their teacher-rated hyperactivity and peer-rated sadness and withdrawal levels were lower, while teacher-rated behavioral engagement increased. Not only did retained children experience teacher-rated behavioral gains, but they also obtained short-term gains in terms of peer-rated liking and sense of school belonging compared to their promoted peers. Similarly, Hong and Yu (2008) found no evidence suggesting that retention at kindergarten is detrimental to children's social-emotional development using PSM. Furthermore, they argue that promotion to first grade might bring long-term harm to children's self-confidence and interest in reading and all school subjects. It should be noted that positive

effect of retention on social outcome tends to be limited to studies that focus on the younger populations, and a lack of maturity is one of the most cited reasons by teachers to retain a child. As a result, it makes sense that if granted more time, younger children who are retained may have additional time to further develop their self-regulation abilities such as to pay attention, inhibit motor activity, persist on tasks, and manage their emotions (Blair, 2002; Kochanska, Murray, & Coy, 1997). In other words, retention may provide children with the opportunity to catch up with their peers in behavioral and emotional adjustment (Wu et al., 2010).

THE PRESENT STUDY

Considering the limited studies on grade retention for students with disabilities, as well as the inconclusive findings associated with retention effects on student academic and social outcomes, more research is warranted on the prevalence and impact of grade retention for students with disabilities using nationally representative data. There are two main motivations underlying the current study of grade retention specifically among students with disabilities. First, students with disabilities are more likely to be at risk for grade retention, but it remains unclear whether there are benefits to repeat a grade, and how retention impacts academic or social outcomes. Therefore, it is important to examine what percent of students with disabilities are retained and to what extent is retention an effective intervention to improve their academic outcomes. Given the existing findings of retention rates, it is equally important to further investigate whether the retention rates among

students with disabilities differ in student characteristics, such as grade level, gender, race, family socioeconomic status, disability type, and maternal education level. In addition, this information can inform budget decisions given that retention is a costly intervention for schools. Bowman (2005) revealed that a school with budget of $6,000 per student would spend $90,000 if it retained 15 students. For students with disabilities, the financial cost associated with retention could be even higher when combined with the costs of special education services. Findings of positive retention effects for students with disabilities could help justify the cost and further inform policy makers of appropriate remedial strategies and interventions for students with disabilities.

In summary, this study examines the retention rate and impact of retention on students with disabilities using data from the *Special Education Elementary Longitudinal Study* (SEELS), a large nationally representative sample of students with disabilities. This study examines the retention rate among students with disabilities overall and by student characteristics such as race, gender, disability type, grade level, and other socioeconomic and cognitive indicators. In addition, the study explores the academic and behavioral trajectories of students with disabilities grouped by retention status at kindergarten, first and second grade. The study further investigates the causal effect of grade retention using propensity score matching method to control for student demographics and other cognitive or behavioral measures so that biases conditional on observable characteristics can be reduced. Three research questions were addressed in this study:

1. What is the rate of grade retention for students with disabilities in the elementary grades, and does retention rate vary by student demographic characteristics?

2. What are the academic and social-behavioral trajectories for students with disabilities retained at kindergarten, first, and second grade?

3. Does retention have a causal effect on the academic outcomes for student with disabilities?

I hypothesize that grade retention rate for students with disabilities will be higher than that of students without disabilities and the rate will differ across student characteristics among students with disabilities. Similar to what was found among students without disabilities, I would expect the retention rate of students with disabilities to differ by grade level, gender, race, maternal education, and family socioeconomic status considering the general effect of these demographics on retention rates. The academic trajectories of retained students with disabilities as reflected by students' academic performance over a longitudinal span are expected to be better than their promoted peers. In addition, the academic benefits resulting from retention are expected to last longer given students with disabilities are at more risk for school failure and in more need of academic support. In terms of social trajectories, grade retention is expected to benefit younger children's social outcome in short period of time but may show trends of decline in the long term.

CHAPTER 3: METHODOLOGY

DATA SOURCE

I examined the *Special Education Elementary Longitudinal Study* (SEELS) database—a longitudinal, federally funded data set collected and managed by the Stanford Research Institute (SRI). SEELS data were collected between the 1999-2000 and 2004-2005 school years, covering three Waves or data collection time of a nationally representative sample of students (Bowman-Perrott et al., 2013). Students were 7 to 14 years old in Wave 1, 8 to 15 years in Wave 2, and 10 to 17 years in Wave 3. Data were collected from the following sources at each Wave: parent interviews, teacher and school administrator questionnaires, and direct student assessments.

SEELS used a two-stage sampling process to select participants so that the sample could be nationally representative. In the first stage, a stratified random sample of 1,124 local education agencies (LEA) based on some key factors such as geographic region of the county, size, and wealth were selected to enhance representativeness. The second stage involved the stratification of student sample from a total of 245 LEAs and 30 special schools that agreed to participate in the study. Sampled districts were then asked to provided rosters of students enrolled in special education born between 09/01/1986 and 08/31/1993. A stratified random sample of 11,512 students was then selected to participate in the study from those schools (Wagner, Newman, Cameto, Levine, & Garza, 2006).

The data collection method involved in SEELS was divided into three parts. First, telephone interviews were conducted with parents. These interviews had an 85% response rate in Wave 1, 75% response rate in Wave 2, and 74% response rate in Wave 3. Second, the school staff questionnaire was carried out using mailed surveys with a 60% response rate in Wave 1, 59% in Wave 2, and 72% in Wave 3. Last, students' academic abilities were tested using direct assessment. Student assessments had a 63% response rate in Wave 1, 74% in Wave 2, and 82% in Wave 3 (Wei, Blackorby, & Schiller, 2011). The collected data are all at the student-level, with data from parents, teachers, and school administrators linked by the student identifier.

ANALYTIC SAMPLE

The analytic sample for the present study consisted of the full SEELS dataset, as well as a subsample of the SEELS dataset. All students with disabilities in the Public Use SEELS database for whom data on student characteristics were available range in ages from 7 to 17 years old are included. Full sample was used when answering the first and second research questions. In order to answer the third research question, students who met the following criteria were included in the sample: (1) complete data from the parent interview regarding student race, gender, age, disability types, English language learner status defined by primary language spoken at home, suspension history, and family income; (2) complete data from standardized assessment score for reading and math WJ-III subtest scores, as well as the alternative tests measuring functional, social, living, self-

care, and mental skills from the initial Wave; and (3) non-missing data on grade level at the initial and second Wave as well as non-missing data on retention status at second, third, fourth, and fifth grade at the second Wave. There were 3,761, 4,720 and 5,035 students who took the WJ-III Letter-Word Identification subtest; 3,834, 5,108 and 4,722 students who took the WJ-III passage comprehension subtest; 3,729, 4,963, and 4,650 students who took the WJ-III applied problem subtest; and 3,568, 4,946 and 4,627 students who took the WJ-III calculation subtest at Wave 1, Wave 2, and Wave 3 respectively. The final analytical sample to answer the third research question varies depending on the grade level of retention and excluded students who did not have WJ-III subtests scores across all three Waves or those who had missing data on student-level predictors. By applying those limiters, the final analytical sample for retention status at second, third, fourth and fifth grade is 139, 107, 105, and 133 respectively. These students were then lumped together to produce the estimate of retention coefficient in the hierarchical linear model.

MEASURES

Demographic Measures

Student disability category, gender, race, age, English language learner, free lunch status, family income, and geography are included as demographic measures for students. A total of 12 disability types under the Individuals with Disabilities Education Act (IDEA) were obtained from school district rosters when the sample was drawn in 1999: learning disabilities, speech impairments, intellectual disability, emotional disturbances, hearing

impairments, visual impairments, other health impairments, orthopedic impairments, traumatic brain injury, autism, deaf/blindness, and multiple disabilities. Besides disability category, phone interviews with parents had been conducted to collect information about other student demographic background on home language, whether students were placed in inclusive classrooms, and extracurricular experiences, social skills, and behavioral problems. Family characteristics, such as mother education and household income, were also included. Student demographic variables were used as the baseline measures to estimate propensity scores for children to control for the pre-treatment (Wave 1) differences.

Academic Measures

Both reading and math achievement were assessed using research editions of subsets of the Woodcock-Johnson Tests of Achievement III (WJ III) (Woodcock, McGrew, & Mather, 2001a) developed for use in SEELS. Two subtests of math assessment were used: applied problems and calculations. Applied problems measure students' ability to comprehend orally presented test items along with their visual stimulus of text, numbers, and graphs; identify relevant information; and select and conduct calculations to arrive at the correct response. Testing items such as identifying numbers, reading a clock, counting objects, adding or subtracting numbers, multiplication, division, and combinations of these basic operations were included for the measure of applied problems. Calculation measures a student's computation skills using a wide range of problems from basic math operations

to geometric, trigonometric, logarithmic, matrix, and calculus (Wei, Lenz, & Blackorby, 2013). In terms of reading measures, two subtests of the WJ III were used to assess student reading achievement (Woodcock, McGrew, & Mather, 2001a). One is the WJ III Letter Word Identification (LWI), which measures letter and word identification skills and the other one is WJ III Passage Comprehension (PC), which measures language comprehension and reading skills using a cloze procedure (Wei, Blackorby, & Schiller, 2011).

Test-retest reliabilities for math were reported to be .93 for the applied problems subtest and .86 for the calculation subtest (Schrank, McGrew, & Woodcock, 2001). Test-retest reliabilities for reading were reported to be .85 for LWI and .76 for PC respectively (Schrank, McGrew, & Woodcock, 2001). The WJ III research versions also produce raw scores, W scores, age equivalents, grade equivalents, and normal curve equivalents. Since W scores are presented on an equal-interval growth scale, which is appropriate for examining student growth. I used W scores for all the descriptive and growth curve analyses in this study.

Social-Behavioral Measures

In SEELS, there are a list of teacher-evaluated survey items measuring students' social-behavioral domains. In the teacher survey, teachers were interviewed on a list of questions, which were further grouped into several broader categories. One of the categories is about student's performance and family support. Within the category, there is

a question (C3) developed to ask for teachers' ratings on a total of 15 items covering students' social and behavioral performances such as self-regulation skills, attention, and peer relationship. The 15 items are measuring how often the student has each of the following behaviors in the class: 1. easily transition from one classroom activity to another, 2. follow your (teacher's) directions, 3. ask for what he or she needs in order to do his or her best in class, 4. join group activities without being told to, 5. make friends easily, 6. start conversations rather than waiting for others to talk first, 7. invite others to join activities, 8. control his or her temper in conflict situations with other students, 9. receive criticism well, 10. cooperate with peers without prompting, 11. respond appropriately when pushed or hit by another student, 12. perform up to his or her ability, 13. keep at a task until he or she is finished, even if it takes a long time, 14. complete homework on time, and 15. communicate his or her thoughts and ideas. Teachers' responses were captured using a three-point scale: never, sometimes and very often with large values indicate better social-behavioral performance whereas smaller values indicate not so well behaved in the social-behavioral aspect.

In order to include those survey items into analysis, a standardized composite social-behavioral score was created using the total 15 items from each Wave. Stata command—*alpha*—was used to generate the composite score with a mean of 0 and standard deviation of 1. At Wave 1, the item-test correlations for all 15 items are roughly the same, ranging from 0.51 to 0.69. The scale reliability coefficient is 0.89. At Wave 2, the scale reliability coefficient for the total 15 items is 0.89 and at Wave 3, the scale

reliability coefficient is 0.88. The composite social-behavioral performance at Wave one ranges from -2.01 to 1.56 (M = -0.001, SD = 0.63). At Wave two, the composite social-behavioral performance ranges from -2.02 to 1.20 (M = -0.003, SD = 0.62), and at Wave 3, the composite social-behavioral performance ranges from -2.52 to 1.21 (M = -0.002, SD = 0.62). The newly created standardized composite social-behavioral score was used as one of the outcomes for the later hierarchical linear modeling.

Alternative Assessment

Students' functional skills were measured in direct assessment using the Scale of Independent Behavior-Revised (SIBR; Bruininks, Woodcock, Weatherman, & Hill, 1996). The assessment provides a comprehensive assessment of 14 areas of adaptive behavior and 8 areas of problem behavior. SEELS incorporates four domains of skill measurement from the assessment: functional skills, daily living skills, functional mental skill, and self-care abilities. The alternative assessment was also used as the baseline measures to estimate propensity scores for children to control for the pre-treatment (Wave 1) differences.

Retention Status

Information about grade retention was obtained from the student characteristics survey questionnaire. Parents were asked if the child "through the past school year, has he/she ever been held back a grade in school?" (question D17A), and the if the answer was affirmative, then the responders are required to provide the grade level in the subsequent

question which describes as "What grade was he/she held back?" (question D17B). Using a series of grade level options, the responders provide information on which grade did the child repeat ranging from kindergarten through 12th grade. The focus of this study is the retrospective retention status captured at three grade levels: kindergarten, first, and second grade.

SAMPLING WEIGHTS ADJUSTMENT

Weights were used to produce population estimates. In SEELS, each instrument within a Wave of data collection has an associated weight for cross-instrument or subgroup analysis. The SEELS public-use dataset included replicate survey weights. Each student was assigned weights to ensure accurate representation within each instrument. This includes a teacher survey weight, parent survey weight, and direct assessment weight. The child-specific weight appropriate for analyzing student data was used to adjust for biased estimates of standard errors. However, no longitudinal sampling weight was provided by the data collector or releaser to take into account the longitudinal data structure. It was recommended by SEELS documentations that when two or more sources of data are combined in one analysis, the weights from the Wave with the smallest number of students should be used (Godard et al., 2007). Accordingly, weights from the direct assessment at Wave 1 were applied in growth curve models to address sample selection bias in the current study.

ANALYSIS METHOD

In order to answer the first research question of the retention rate among students with disabilities, I used descriptive statistics to report the frequency and proportion of students with disabilities who were ever retained at kindergarten through fifth grade respectively. I also presented the rates by student and family characteristics. In order to test whether the retention rate is statistically different across student characteristics, I used Wald test differences in means and report significance *p* value associated with the F value test estimates.

In order to answer the second question of what are the academic and social-behavioral trajectories of students ever retained at kindergarten, first and second grade, I used Hierarchical Linear Modeling (HLM; cite) using student reading, math and standardized social-behavioral performance as outcomes separately (dependent variable) while adding retention status at kindergarten through second grade as predictor (independent variable), along with a couple of other student demographic variables such as disability type, student race, gender, EL status, suspension history, free or reduced-price lunch enrollment, and home geography. The estimated regression coefficients and *p* values associated with retention status and all the other independent variables were reported. In addition, trajectories of student academic and social-behavioral outcome from age 7 to 17 years were plotted by retention status at kindergarten through second grade separately using model results to indicate group differences longitudinally. I employed multilevel

model predicting academic (reading and math) outcomes. Previous studies with the SEELS had suggested the superiority of the quadratic growth curve model over the linear model in estimating student growth because the greatest growth had been found in early grades, with a decrease in growth rates as a student becomes older (e.g., Bloom, Hill, Black, & Lipsey, 2008; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Morgan, Farkas, & Hibel, 2008). Therefore, a quadratic growth curve model was used to display the academic and behavioral growth trajectories for students with disabilities from ages 7 to 17. Nevertheless, both a linear model and a quadratic model were fitted to estimate the growth trajectories of students with disabilities and the likelihood ratio test result were presented to compare model fit.

HLM (Raudenbush & Bryk, 2002) was conducted to take into account the nesting of observations within each individual's data. HLM is especially useful for studying individual change over time and how individual characteristics are associated with the average level of achievement or with change of achievement over time. HLM does not assume the spacing of observations to be consistent across persons or Waves, making it suitable for the SEELS data structure given the 1-year gap between Wave 1 and Wave 2 and a 2-year gap between Wave 2 and Wave 3 (Wei, Blackorby & Schiller, 2011). It also has the advantages of minimizing bias due to missing data because it uses all available data points from each person to estimate the growth curves. A two-level HLM with appropriate sampling weights was used for the analyses in this study. Since multilevel modeling does not work with the replicate weights, the HLM model clustered at school level to account

for estimate bias on the effects. The Level 1 HLM model is the within-person model, which included repeated measures of student's academic scores and behavioral measures across three Waves predicted by student's centered age and centered age squared at specified grade level at each Wave. The Level 2 model is the between-person model, which estimated the differences in academic scores and behavioral ratings between students using retention status, disability categories, race, ELL status, sex, free lunch status, suspension history, age at initial wave, baseline (initial wave) academic performances by subtests (letter word identification, passage comprehension, applied problems and calculation), and family location as covariates. Intercept, age, and age squared were modeled as random effects. Retention status, disability types, race, ELL status, sex, free lunch status, suspension history, age at initial wave, baseline academic performances by subtest, and family location were modeled as fixed effects. Restricted maximum likelihood estimation with standard error robust option cluster at the school level was specified. The two-level HLM models used are as follows:

Level 1:

$$Y_{ti} = \beta_{0i} + \beta_{1i}(Age_{ti} - 12.67) + \beta_{2i}(Age_{ti} - 12.67)^2 + \varepsilon_{ti}$$

in which Y_{ti} is academic or behavioral scores of student *i* in time t, β_{0i} is the average level of student *i* at age 12.67, β_{1i} is the linear slope of student *i* at age 12.67, β_{2i} is the quadratic curvature of acceleration of growth for student *i* over time, and ε_{ti} is the residual term. Since age was centered by the mean of 12.67, so β_{0i} becomes meaningful, representing the academic or behavioral scores at 12.67 years, instead of at 0 years old if age was not

centered (Raudenbush & Bryk, 2002). β_{1i} describes the slope at age 12.67. β_{2i} expresses the change in slope, or acceleration over time. The level, slope, and acceleration in Level 1 model are the outcomes for the Level 2 models:

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(Retain_K) + \gamma_{02}(Disability) + \gamma_{03}(Race) + \gamma_{04}(EL) + \gamma_{05}(Sex) + \gamma_{06}(Suspension) + \gamma_{07}(FRL) + \gamma_{08}(Location) + \gamma_{09}(age\ at\ initial\ wave) + \gamma_{10}(academic\ performance\ by\ subtests\ at\ initial\ wave) + r_{0j}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(Retain_K) + r_{1j}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}(Retain_K) + r_{2j}$$

Retention at kindergarten (and other grades) is a binary variable using the promoted group as the reference group. Disability refers to a set of dummy variables with a separate dummy variable representing a different disability group versus the reference group (students with autism). Race also refers to a set of dummy variables with each dummy variable representing a racial group versus the reference racial group (White). EL status is a binary variable using non-ELs as the reference group. Sex, suspension, and free lunch status are represented by binary variables with females, those who were never suspended, and not enrolled in free lunch program to be the reference group. Family location is a categorical variable with urban as the reference group. Student age at the beginning of grade and baseline academic performances by subtests were extracted from Wave 1. γ_{00}, γ_{10}, and γ_{20} represent the average academic or behavioral achievement, growth rate, and

acceleration of academic or behavioral growth for 12.67-year-old White female, non-EL students with autism, not enrolled in free lunch program, was never suspended, and living in the urban area who were promoted to next grade level instead of being retained. To visually illustrate what the academic and behavioral trajectories of students retained at kindergarten are like, plots of academic and behavioral performances by retention status from age 7 to 17 were displayed. In order to examine whether the effect of retention at kindergarten on student academic and social-behavioral outcomes changes over time, I tested the interactions between retention status at kindergarten with age and age squared. A significant interaction indicates the effects of retention at kindergarten change over time. To further examine academic and behavioral trajectories of students retained at first and second grade, I repeated the steps described above and switched the retention status from kindergarten to first and second grade separately and generate the corresponding trajectories by retention status at first and second grade respectively.

In order to answer the last research question of how retention status from second to fifth grade impacts academic trajectories for students with disabilities, I first identified four groups of students based on retention status at second through fifth grade. Group one represents retention at second grade group and it includes students who (1) were at second grade at the initial Wave, (2) had non-missing retention status at second grade at Wave 2 (0 or 1), and (3) had non-missing values in all baseline variables. Group two represents retention at third grade group and it includes students who (1) were at third grade at the initial Wave, (2) had non-missing retention status at third grade at Wave 2 and (3) had non-

missing values in all baseline variables. Group three represents retention at fourth grade group and it includes students who (1) were at fourth grade at the initial Wave, (2) had non-missing retention status at fourth grade at Wave 2, and (3) had non-missing values in all baseline variables. Group four represents retention at fifth grade and it includes students who (1) were at fifth grade at the initial Wave, (2) had non-missing retention status at fifth grade at Wave 2, and (3) had non-missing values in all baseline variables. Retention status at kindergarten and first grade were not included because of small sample size and propensity score matching thus cannot be properly implemented. In doing so, I was able to include the initial Wave measurements as the baseline measure when estimating the propensity scores for students retained at second through fifth grade. Next, I used propensity score matching (PSM) to create two statistically equivalent groups (promoted vs. retained) based on a set of baseline variables for each of the four groups. PSM assumes the strong ignorability assumption in which the treatment assignment is not associated with unmeasured covariates given the observed covariates and it's especially effective if a comprehensive list of pretreatment covariates were collected and controlled for (Rosenbaum & Rubin, 1983; Rosenbaum, 2002). The propensity score is a one-dimensional balancing score defined as the probability of treatment given a set of covariates that theoretically have impact on the treatment (Rosenbaum & Rubin, 1983). With the strong ignorability assumption holds, adjusting the propensity scores is sufficient for an unbiased estimation of causal effects. Propensity scores, the predicted probability of being retained in kindergarten, first, or second grade separately, were estimated for the full

analytical sample of children for whom retention and other demographic information was available. In order to obtain the maximized number of matched cases using observations from the retained and promoted participants, a total of 15 background variables collected at the initial Wave were used, including child demographic variables. Lastly, students' prior social skills measured using the alternative assessments in the initial Wave, such as functional skill, social skill, independent living skill, mental skill and self-care skills were also included in estimating propensity scores.

A logistic regression model was used to estimate propensity scores (Rosenbaum, 2010; Rosenbaum & Rubin, 1983):

$$ln\frac{\hat{p}}{1-\hat{p}} = b_0 + \sum_{i=0}^{i=15} b_i X_i$$

where $\hat{p}$ is the estimated probability of being in the retained group, X_i is the i^{th} baseline predictor, b_0 is the intercept, and b_i is the regression coefficient for the ith predictor. The term on the left side of the equation can be transformed to a probability of being retained conditional on the student's level on the baseline variables.

Using the estimated propensity score, treatment cases were matched to control cases using the nearest available pair matching with replacement. The order of treatment and control cases was random and for each treatment case, a control case with the smallest absolute difference in propensity score was then selected. This procedure continues until all treatment cases have a match. Stata 14.0 (College Station, TX) was used to implement the matching and the command used to implement matching procedure is an official Stata package named *teffects psmatch*. The command implements a variety of propensity score

matching methods to adjust for pre-treatment observable differences between a group of treated and a group of control by calculating approximate standard errors on the treatment effects, assuming independent observations and homoscedasticity of the outcome variable within the treated and control groups. As a result, the variance of the outcome does not depend on the propensity score. Treatment status is identified by four variables representing retention status at second, third, fourth and fifth grade separately. The Stata command offers various matching methods, such as one-to-one, k-nearest neighbors, radius, kernel, local linear regression and Mahalanobis matching. For the current analysis, a default one-to-many matching with a caliper value of 0.2 was imposed. Caliper value dictates the maximum distance in propensity scores allowed for a treated unit to be matched with untreated units. By imposing caliper value to be 0.2, any pair of retained and promoted children who differed in their propensity scores by more than 0.2 could not be matched with each other. The command automatically generates a new variable containing student identification numbers of those who were successfully matched, which will later be used in the growth curve model. I repeated the matching process using second, third, fourth and fifth grade retention variables separately to generate different sets of matched cases. In order to check balance among the matched cases, I reported the standardized differences for each of the predictors used to estimate propensity scores before and after the matching procedure for each grade level. In addition, I also included density plots of propensity scores before and after matching.

After matching the retained students with their promoted peers, I stacked matched students by the retained and promoted group and applied hierarchical linear regression (HLM) longitudinal growth modeling to examine the academic trajectories over time for both groups and compare whether the retention group over-perform the promoted group. The two-level HLM models used are as follows:

Level 1:

$$Y_{ti} = \beta_{0i} + \beta_{1i}(Age_{ti} - 11.25) + \beta_{2i}(Age_{ti} - 11.25)^2 + \varepsilon_{ti}$$

in which Y_{ti} is academic or behavioral scores of student *i* in time t, β_{0i} is the average level of student *i* at age 11.25, β_{1i} is the linear slope of student *i* at age 11.25, π_{2i} is the quadratic curvature of acceleration of growth for student *i* over time, and ε_{ti} is the residual term. Since age was centered by the mean 11.25, so β_{0i} becomes meaningful, representing the academic or behavioral scores at 11.25 years, instead of at 0 years old if age was not centered (Raudenbush & Bryk, 2002). β_{1i} describes the slope at age 11.25. β_{2i} expresses the change in slope, or acceleration over time. The level, slope and acceleration in Level 1 model are the outcomes for the Level 2 models:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(Retained) + \gamma_{02}(Disability) + \gamma_{03}(Race) + \gamma_{04}(EL) + r_{0j}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(Retained) + r_{1j}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}(Retained) + r_{2j}$$

Retention is a binary variable using the promoted group as the reference group. Disability refers to a set of dummy variables with a separate dummy variable representing a different

disability group versus the reference group (students with Autism). Race also refers to a set of dummy variables with each dummy variable representing a racial group versus the reference racial group (white). ELL status is a binary variable using non-ELs as the reference group. β_{00}, β_{10}, and β_{20} represent the average academic achievement, growth rate, and acceleration of academic growth for 11.25-year-old White, non-EL students with autism who were promoted instead of retained.

CHAPTER 4: RESULTS

DESCRIPTIVE ANALYSIS ON RETENTION RATE AMONG STUDENTS WITH DISABILITIES

Descriptive statistics regarding retention rates among students with disabilities are presented in table 1. Using the retention status variable from kindergarten to fifth grade, I presented the percent of students who were ever retained and never retained from kindergarten to fifth grade. About 92% students with disabilities were ever retained from kindergarten to fifth grade whereas only 8% students with disabilities were never retained. Risk rates of being ever retained from kindergarten to fifth grade disaggregated by student demographics such as sex, race, disability types, language status, free lunch program enrollment, suspension status, maternal education and family geography were also reported. For instance, retention rates are higher for students who are female (93%), Hispanic (93%), have specific language impairment (96%), English language learner (93%), not enrolled in the free lunch program (91%), never suspended (95%), have mother graduated with high school degree (92%), and live in the rural areas (92%). In addition to the overall retention status across the six primary grades (kindergarten through fifth grade), the risk rates by student demographics disaggregated on the frequency of retention from kindergarten to fifth grade was also presented. Among all students in primary grades, the majority of them were retained twice (44%), followed by those retained once (37%), three times (9.2%) and four or five times (2.5%). Female students consistently are at higher risk of being retained than male students, regardless of the frequency. In terms of other student demographic variables, the patterns of risk rates of being retained by frequency vary although the risk rates may not differ as much. For instance, students with other health impairments are having higher rates of being retained twice (49%) whereas students with orthopedic impairment are having higher rates of being retained three times (12%). English

language learners are having higher rates of being retained once (40%) and twice (44%) whereas English native speakers (8.7%) are having higher rates of being retained three times. Speaking of free lunch enrollment, students not enrolled in free lunch programs are having higher rates of being retained once (41%) whereas students enrolled in free lunch program are having higher rates of being retained twice times (44%), three times (11%) and four or more times (3%). Students who were never suspended have higher rates of being retained once (45%), three times (8.6%) and four or more (3.1%) whereas students who were ever suspended have higher rates of being retained twice (41%). Rates of being retained twice (44%) and three times (10%) are higher for students who have mothers graduated with high school degrees whereas rates of being retained four or more (2%) are higher for students who have mothers graduated with less than high school degrees. In terms of family geography, rates of being retained twice (42%) are higher for students living in the urban areas whereas rates of being retained three times are higher for students living in the rural areas (11%).

	Ever retained from grades K-5	Frequency of retention from K to grade 5			
		Retained once	Retained twice	Retained 3 times	Retained 4 or more times
Proportion	0.924	0.365	0.442	0.092	0.025
Gender					
Male	0.896	0.378	0.379	0.073	0.008
Female	0.927	0.424	0.427	0.093	0.031
Race					
White	0.914	0.395	0.435	0.076	0.008
Black	0.877	0.412	0.327	0.107	0.031
Hispanic	0.925	0.412	0.439	0.056	0.019
Asian	0.921	0.392	0.286	0.226	0.018
Other	0.910	0.483	0.302	0.124	--
Disability					
SLD	0.904	0.391	0.419	0.084	0.009
SLI	0.957	0.415	0.433	0.097	0.013
MR	0.921	0.464	0.334	0.095	0.028
EBD	0.752	0.415	0.251	0.058	0.027
HI	0.902	0.445	0.375	0.061	0.021
VI	0.915	0.460	0.378	0.065	0.013
OI	0.895	0.342	0.411	0.117	0.025
OHI	0.875	0.294	0.489	0.065	0.026
AUT	0.857	0.364	0.375	0.084	0.034
OLI	0.907	0.401	0.403	0.069	0.034
Language status					
Non-EL	0.901	0.401	0.399	0.087	0.015
EL	0.933	0.403	0.437	0.078	0.015
Free lunch					
Non-FRPL	0.907	0.410	0.408	0.083	0.007
FRPL	0.903	0.333	0.436	0.105	0.029
Suspension status					
Never suspended	0.949	0.445	0.401	0.086	0.031
Ever suspended	0.819	0.308	0.413	0.083	0.012
Maternal education					
Less than HS	0.912	0.410	0.387	0.094	0.021
HS	0.920	0.376	0.437	0.099	0.010
Some college	0.891	0.418	0.408	0.055	0.011
Bachelor or higher	0.915	0.484	0.364	0.057	0.009
Geography					
Rural	0.923	0.413	0.396	0.107	0.006
Suburban	0.902	0.373	0.422	0.089	0.018
Urban	0.907	0.421	0.389	0.078	0.018

Table 1: Risk Rates of Students Ever Retained from Kindergarten to Grade 5 (n=13,176)

Note. Estimates include individual-level replicate survey weights to account for complex survey design. Other race includes American Indian or Alaska native and multi or other races. Other low incidence disabilities include students with multiple disabilities and those who are deaf or blind.

Table 1, cont.

Child demographics	Kindergarten			Grade 1			Grade 2		
	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test
Proportion	0.264	0.736		0.385	0.615		0.426	0.574	
Male	0.691	0.663	0.27	0.643	0.686	1.09	0.632	0.689	0.13
Race									
White	0.792	0.599	3.71	0.638	0.659	0.00	0.659	0.647	0.11
Black	0.131	0.225	2.05	0.172	0.216	0.76	0.222	0.189	0.06
Hispanic	0.042	0.151	4.35	0.158	0.100	0.91	0.106	0.129	1.27
Asian	0.015	0.019	0.10	0.018	0.017	0.07	0.012	0.020	0.56
Other	0.019	0.007	0.14	0.013	0.009	0.00	0.001	0.015	0.12
Disability									
SLD	0.497	0.471	0.60	0.531	0.418	1.12	0.489	0.489	1.46
SLI	0.195	0.283	0.00	0.230	0.268	0.00	0.332	0.215	2.96
MR	0.142	0.087	2.06	0.105	0.128	0.82	0.108	0.126	0.05
EBD	0.034	0.050	1.37	0.033	0.050	0.41	0.037	0.048	2.52
HI	0.014	0.016	0.05	0.015	0.015	0.00	0.013	0.016	0.46
VI	0.003	0.003	0.42	0.004	0.005	0.12	0.004	0.005	0.82
OI	0.020	0.010	0.34	0.013	0.014	0.01	0.016	0.013	0.16
OHI	0.056	0.055	0.03	0.038	0.060	2.94	0.047	0.054	0.02
AUT	0.016	0.016	0.62	0.014	0.017	0.32	0.016	0.016	0.00
Other low incidence	0.022	0.010	3.29	0.014	0.025	5.32*	0.027	0.019	0.97
EL	0.027	0.159	3.30	0.153	0.131	0.00	0.141	0.138	0.08
FRPL	0.553	0.592	1.41	0.574	0.585	0.05	0.618	0.562	0.30

Table 2: Composition Rates of Retention Status at Kindergarten through Grade 2 by Student Demographic (n=13,176)

	Kindergarten			Grade 1			Grade 2		
	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test
Maternal education									
Less than HS	0.207	0.229	1.00	0.206	0.218	0.02	0.228	0.206	0.03
HS	0.432	0.434	0.09	0.502	0.403	1.04	0.419	0.449	0.13
Some college	0.227	0.208	0.21	0.209	0.259	0.24	0.254	0.234	0.05
Bachelor or higher	0.133	0.130	0.76	0.083	0.120	0.72	0.099	0.111	0.04
Geography									
Rural	0.269	0.162	1.64	0.214	0.161	0.56	0.176	0.182	0.00
Suburban	0.538	0.487	1.08	0.452	0.541	0.00	0.525	0.500	0.03
Urban	0.183	0.351	3.22	0.333	0.298	0.47	0.299	0.317	0.03
Ever suspended	0.380	0.367	0.04	0.200	0.362	5.60*	0.332	0.287	0.02
Academic outcome									
Reading									
LWI	472.52	476.37	0.030	470.54	477.97	3.360	473.95	476.26	0.280
PC	482.20	484.29	0.050	482.13	484.62	2.190	482.60	484.45	0.210
Math									
AP	486.05	490.26	0.460	487.18	490.26	0.570	485.67	491.29	0.640
Cal	496.51	500.53	1.670	497.02	500.83	0.560	497.24	500.86	1.630
Social-behavioral									
Composite score	0.075	0.068	0.040	0.068	0.071	0.000	0.113	0.045	0.310

Note. Wald test indicates whether there are significant differences in means and proportions for retained and promoted groups by student demographic and outcome measure at each grade level. Binary and categorical variables are the Wald tests from the logistic regression. Continuous variables are the F tests from the analyses of variances. *p< .05, **p<.01, ***p<.001

Table 2, cont.

Child demographics	Grade3			Grade 4			Grade 5		
	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test	Retained	Promoted	*Wald* χ^2/ F test
Proportion	0.195	0.805		0.124	0.876		0.070	0.930	
Male	0.647	0.683	0.00	0.643	0.675	0.02	0.638	0.674	0.09
Race									
White	0.607	0.675	0.53	0.520	0.676	1.82	0.613	0.655	0.60
Black	0.197	0.202	0.04	0.320	0.178	1.79	0.223	0.197	0.39
Hispanic	0.179	0.089	2.02	0.157	0.115	0.00	0.136	0.120	0.08
Asian	0.011	0.021	0.18	0.001	0.020	0.04	0.028	0.016	0.01
Other	0.006	0.013	0.01	0.001	0.012	1.41	0.000	0.011	0.06
Disability									
SLD	0.420	0.420	0.49	0.461	0.461	0.00	0.462	0.462	0.03
SLI	0.227	0.269	0.00	0.224	0.259	0.01	0.165	0.264	0.28
MR	0.098	0.131	1.57	0.156	0.113	0.54	0.128	0.119	0.02
EBD	0.027	0.054	2.13	0.046	0.044	0.07	0.059	0.043	0.09
HI	0.013	0.016	0.20	0.015	0.015	0.06	0.014	0.015	0.03
VI	0.004	0.005	0.16	0.003	0.005	1.84	0.007	0.004	0.81
OI	0.012	0.015	0.00	0.014	0.014	0.26	0.017	0.013	0.11
OHI	0.059	0.048	0.55	0.059	0.050	0.00	0.121	0.044	2.15
AUT	0.012	0.018	0.60	0.015	0.017	0.70	0.029	0.015	0.64
Other low incidence	0.016	0.024	1.93	0.018	0.022	0.00	0.026	0.021	0.00
EL	0.197	0.108	1.11	0.199	0.128	1.02	0.203	0.132	0.96
FRPL	0.629	0.555	1.39	0.669	0.564	2.00	0.624	0.576	0.18

Table 3. Composite Rates of Retention Status at Grade 3 through 5 by Student Demographics (n=13,176)

	Grade 3			Grade 4			Grade 5		
	Retained	Promoted		Retained	Promoted		Retained	Promoted	
Maternal education									
Less than HS	0.236	0.201	0.62	0.263	0.204	0.05	0.197	0.215	0.62
HS	0.464	0.425	0.00	0.383	0.450	0.04	0.486	0.434	0.12
Some college	0.222	0.251	0.17	0.263	0.237	0.04	0.223	0.243	0.58
Bachelor or higher	0.078	0.122	0.99	0.092	0.110	0.07	0.094	0.108	0.01
Geography									
Rural	0.160	0.192	0.08	0.088	0.198	1.54	0.088	0.190	6.53
Suburban	0.506	0.510	0.00	0.474	0.515	0.48	0.579	0.501	0.00
Urban	0.334	0.299	0.07	0.438	0.288	2.35	0.334	0.309	0.28
Ever suspended	0.262	0.324	0.12	0.415	0.281	0.27	0.372	0.294	0.74
Academic outcome									
Reading									
LWI	474.75	474.31	0.070	476.04	474.18	0.020	480.56	473.92	0.340
PC	480.95	483.32	0.720	482.41	482.94	0.080	484.51	482.75	0.010
Math									
AP	488.10	487.76	0.020	489.27	487.63	0.010	491.40	487.55	0.510
Cal	498.78	499.51	0.020	502.37	498.46	0.000	504.38	498.50	1.820
Social-behavioral outcome									
Composite score	0.121	0.077	0.050	0.017	0.094	0.140	-0.069	0.094	0.340

Note. Wald test indicates whether there are significant differences in means and proportions for retained and promoted groups by student demographic and outcome measure at each grade level. Binary and categorical variables are the Wald tests from the logistic regression. Continuous variables are the F tests from the analyses of variances. *p< .05, **p<.01, ***p<.001

Table 3, cont.

Table 2 and 3 present the composition rates of retention disaggregated by grade level from kindergarten to fifth grade using student demographic variables as well as the outcome measures. Among all kindergarteners, a total of 26% students were retained, and the rate goes up to 39% at first grade and 43% at second grade. Starting from third grade, the rate goes down to less than 20% and by fifth grade, less than 10% students were retained. Comparisons of percentage students retained versus promoted by student demographic characteristics were also presented. For instance, among all kindergarteners who were retained, about 50% are students with learning disabilities, a rate comparable to kindergarteners with learning disabilities (47%) who were promoted. Among all retained first graders, about 64% are white, similar to the percentage (66%) of promoted first graders who were white. At fourth grade, about 20% of retained students are English language learners, 7% fewer than the percentage of promoted students who were English language learners. Nevertheless, Wald test results indicate there is no statistical difference for both groups (retained and promoted) in terms of demographic background from kindergarten to fifth grade except for disability type and suspension history at first grade. There are significantly more students with other low incidence disabilities (multiple disabilities or deaf and blind) promoted at first grade than retained (χ^2=5.32, df=1) whereas there are significantly fewer students who were ever suspended being retained at first grade than promoted (χ^2=5.60, df=1). Compared to students who were promoted, student retained at kindergarten have lower scores on all academic measures such as letter word identification, passage comprehension applied problem and calculation. Retained students consistently have lower academic performance than their promoted peers until third and fourth grade when retained students achieved slightly higher on letter word identification, applied problem and calculation than their peers promoted to the next grade. Moreover, students retained at fifth grade have greater performance on all four academic subtests than

their promoted peers. As for social-behavioral outcomes captured using the composite scores, students retained at kindergarten and second grade have higher composite scores than their promoted peers. In contrast, the composite scores for students retained at first, third, fourth and fifth grade are lower than their promoted peers. However, both groups do not differ significantly in terms of academic and social-behavioral outcome as none of the F test results is significant.

ACADEMIC TRAJECTORIES OF STUDENTS EVER RETAINED AT KINDERGARTEN, FIRST, OR SECOND GRADE

Tables 4 through 9 include model results predicting academic outcomes by subject areas and retention status from kindergarten through second grade. Table 4 and 5 present the HLM models for estimating the growth curves in academic outcome measures for students with disabilities from ages 7 to 17 by retention status at kindergarten. Table 6 and 7 present the HLM models for academic outcome measures by retention status at first grade and Table 8 and 9 present the models for academic outcome measures by retention status at second grade. For each outcome, three models were included. Model 1, the base model, contains the linear, quadratic age and retention status at kindergarten as predictors. Model 2 adds interactions between retention status and the linear and quadratic term of age. Model 3 adds all the other covariates and interaction terms if they were significant from model 2. The coefficients of fixed effects indicate the relationship between the fixed-effect predictors and level, slope and acceleration. Standard errors are presented in the parenthesis below the regression coefficients. Figures 1, 2 and 3 present the estimated growth curves on four academic subtests generated from the HLM models for students with disabilities by retention status at kindergarten, first, and second grade, respectively.

Kindergarten

Letter-word identification. In Table 4, the average level was significant and indicated that the average letter word identification scores for the reference group (white non-EL male students, not enrolled in free or reduced-price lunch programs, have autism, never suspended before, live in urban areas, and promoted to next grade level) were 148.64 at age 12.22 years while adjusting for baseline letter word identification performance and student age at the initial wave. Results showed a significant initial linear slope in the trajectory of letter word identification scores for the promoted group, $\gamma_{10} = 7.41, p < .001$, which was modified by a significant negative quadratic effect, $\gamma_{20} = -0.51, p < .001$. This pattern indicated that acceleration rate of yearly gain of letter word identification slightly decreased in each subsequent year. The effect of retention at kindergarten was not significant across three models, indicating retained students do not score significantly different than their peers who were promoted, even after controlling for the effect of all covariates in the full model. Interactions between retention status and student age and age squared were not significant, indicating letter word identification trajectories between two groups do not differ over the 10-year period. Student age at the initial Wave is a negative predictor indicating students who are one year older at the beginning of the grade scored on average 8.29 points significantly lower on letter word identification. In contrast, baseline performance at the initial Wave is a positive predictor indicating students who scored one point higher on letter word identification at the beginning scored on average 0.90 points significantly higher. Besides the significant effects of baseline letter word identification performance and student age at the initial Wave, suspension history also produced significant result. Students who were suspended before scored about 3.37 points significantly lower than students who were never suspended.

Passage comprehension. In the base model predicting passage comprehension performance in table 4, the linear change in passage comprehension was positive and significant, γ_{10} = 4.01, p < .001, indicating that students with disabilities who were promoted instead of being retained at kindergarten grew 4.01 score points on passage comprehension at age 12.22. Similarly, the significant negative acceleration γ_{20} = -0.35 indicated students' growth on passage comprehension was slower as they grow older. The average passage comprehension for the reference group is 155.02 at age 12.22 years while adjusting for baseline passage comprehension performance and student age at the initial Wave. The effect of retention at kindergarten was not significant in all three models, indicating students retained at kindergarten did not perform significantly different in passage comprehension compared to their peers who were promoted even after controlling for all the other covariates. Interactions between retention status and age are nonsignificant, indicating the gaps of passage comprehension trajectories between the two groups do not change over time. The change in the outcome is more related to student age at the initial Wave, baseline passage comprehension performance, disability types and EL status. Students who are one year older at the beginning of the grade scored significantly lower (γ_{09} = 4.47, p<.001) whereas students scored higher on baseline passage comprehension scored significantly higher (γ_{10} = 0.78, p<.001) on the outcome. Compared to students with autism, students with SLD ($\gamma_{02[SLD]}$ = 5.21, p<.001), SLI ($\gamma_{02[SLI]}$ = 6.84, p<.001), EBD ($\gamma_{02[EBD]}$ = 7.20, p<.001), VI ($\gamma_{02[VI]}$ = 5.81, p<.01), OI (($\gamma_{02[OI]}$ = 4.90, p<.001), OHI ($\gamma_{02[OHI]}$ = 4.85, p<.01) scored significantly higher on the outcome. In contrast, students who were suspended before scored 2.06 significantly lower than their peers who were never suspended before.

Applied problem. Table 5 presents the HLM models for estimating growth curves in applied problems for students with disabilities by retention status at kindergarten. The

average level of applied problem for reference group is 135.64 at age 12.22 while adjusting for the effects of baseline performance and student age at the initial Wave. The significant initial slope, γ_{10} = 4.61, p<.001 modified by a significant negative quadratic effect γ_{20} = -0.30, p<.001 indicates as students grow older, the trajectories on applied problem get plateaued. The fixed effect associated with retention status at kindergarten is not significant, and the interactions between retention status at kindergarten and age or age squared are not significant as well, indicating the gaps between the two groups on applied problems are constant over time. Students who are one year older at the beginning of the grade scored 4.28 points significantly lower whereas those who scored one unit higher on baseline applied problem scored 0.83 significantly higher. Besides student age at initial Wave and baseline performance, disability type and family location also produced significant results. For instance, compared to students with autism, students with SLD, SLI, EBD, HI and OHI scored on average 4.02, 6.47, 6.47, 6.09, and 4.68 points significantly higher on the outcome. Students with disabilities who lived in the rural areas scored on average 4.39 points significantly higher than their peers who lived in the urban areas.

Calculation. Table 5 presents the HLM models for estimating growth curves in calculation measurement for students with disabilities by retention status at kindergarten. The significant initial slope, γ_{10} = 4.32, p<.001 modified by a significant negative quadratic effect γ_{20} = -0.47, p<.001 indicates as students grow older, the trajectories on calculation get plateaued. The fixed effect of retention at kindergarten on student calculation performance is not significant while controlling for the effect of baseline performance, student age at the initial Wave, and interactions between retention and student age. Baseline performance is a positive significant predictor while student age at the initial Wave is a significant negative predictor. The significant effect of baseline performance and student age at the initial Wave persist even after controlling for the effects

of all the other covariates. None of the interactions between retention and student age is significant and in the final model where all the other covariates are controlled for, the effect of retention at kindergarten becomes significant and students retained at kindergarten scored on average 2.73 points lower than students who were promoted at kindergarten (γ_{01} = -2.73, p<.05). Besides retention, disability type and home geography also produced significant coefficients. For example, compared to students with autism, students with SLI scored 5.96 points significantly higher on calculation. Students with disabilities living in the rural areas scored on average 4.87 points significantly higher than their peers from urban areas.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
Age	7.409***	7.209***	7.232***	4.010***	3.913***	3.977***
	(0.339)	(0.374)	(0.367)	(0.263)	(0.275)	(0.294)
Age^2	-0.505***	-0.522***	-0.439***	-0.352***	-0.329***	-0.361***
	(0.068)	(0.076)	(0.075)	(0.071)	(0.078)	(0.070)
Age at initial wave	-9.001***	-9.012***	-8.294***	-4.707***	-4.693***	-4.467***
	(0.536)	(0.529)	(0.518)	(0.366)	(0.367)	(0.393)
LW at initial wave	0.916***	0.916***	0.899***	-	-	-
	(0.020)	(0.020)	(0.021)			
PC at initial wave	-	-	-	0.780***	0.780***	0.780***
				(0.025)	(0.025)	(0.029)
Retained at K	1.420	1.232	1.905	0.409	0.977	0.463
	(1.139)	(1.321)	(1.306)	(0.982)	(1.288)	(1.041)
Retained at K # age		0.768			0.383	
		(0.491)			(0.394)	
Retained at K # age^2		0.030			-0.112	
		(0.155)			(0.156)	
SLD			0.006			5.209***
			(2.166)			(1.472)
SLI			0.747			6.836***
			(2.374)			(1.644)

Table 4. Hierarchical Linear Regression Model Predicting Reading Performance (Retained at Kindergarten)

	Letter word identification			Passage comprehension		
	M1	M2	M3	M1	M2	M3
MR			-0.848			3.245
			(2.442)			(1.688)
EBD			2.311			7.196***
			(2.377)			(2.014)
HI			0.490			2.290
			(2.205)			(1.481)
VI			1.309			5.854**
			(2.496)			(1.886)
OI			-0.691			4.897***
			(2.495)			(1.474)
OHI			3.253			4.849**
			(2.464)			(1.522)
OLI			-1.852			1.247
			(2.129)			(1.604)
Black			-0.650			-1.451
			(1.346)			(1.387)
Hispanic			4.005			-1.439
			(2.710)			(1.944)
Asian			0.944			2.402
			(2.056)			(2.013)
Other			-3.182			-9.934
			(2.323)			(5.318)
Male			1.762			0.587
			(1.226)			(0.958)
EL			-1.998			1.856
			(2.915)			(1.792)

Table 4, cont.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M1	M2	M3
Suspension			-3.365**			-2.064*
			(1.121)			(1.008)
FRPL			-0.727			-0.381
			(1.193)			(1.179)
			(1.383)			(1.206)
Suburban			-0.709			1.425
			(1.168)			(1.074)
_cons	148.635***	148.753***	148.421***	163.151***	162.916***	155.023***
	(9.661)	(9.623)	(11.016)	(10.871)	(10.887)	(12.881)
Random Effect						
Student	45.900***	45.246***	31.218***	24.165***	24.024***	15.972***
	(11.623)	(11.287)	(5.857)	(7.233)	(7.196)	(5.840)
Residual	151.801***	151.732***	148.701***	117.505***	117.416***	117.003***
	(14.501)	(14.618)	(15.928)	(13.095)	(13.080)	(14.386)
N	2773	2773	2491	2847	2847	2552
ll	-7838199	-7835428	-6854618	-7566268	-7564853	-6635425

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
Age	4.609***	4.448***	4.651***	4.323***	4.372***	4.399***
	(0.354)	(0.415)	(0.354)	(0.358)	(0.379)	(0.389)
Age^2	-0.295***	-0.321***	-0.344***	-0.465***	-0.435***	-0.480***
	(0.071)	(0.089)	(0.070)	(0.067)	(0.077)	(0.076)
Age at initial wave	-4.282***	-4.299***	-4.481***	-4.202***	-4.184***	-3.718***
	(0.509)	(0.512)	(0.492)	(0.538)	(0.548)	(0.652)
AP at initial wave	0.833***	0.833***	0.817***	-	-	-
	(0.020)	(0.020)	(0.022)			
CAL at initial wave	-	-	-	0.727***	0.727***	0.686***
				(0.042)	(0.043)	(0.045)
Retained at K	-1.740	-2.145	-1.258	-1.916	-1.346	-2.726*
	(1.179)	(1.435)	(1.369)	(1.237)	(1.653)	(1.210)
Retained at K # age		0.612			-0.175	
		(0.558)			(0.410)	
Retained at K # age^2		0.071			-0.108	
		(0.159)			(0.162)	
SLD			4.018*			0.490
			(1.945)			(1.842)
SLI			6.469**			5.961**
			(2.122)			(2.050)

Table 5. Hierarchical Linear Regression Model Predicting Math Performance (Retained at Kindergarten)

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
MR			-0.122			-2.174
			(2.015)			(2.336)
EBD			6.471**			1.727
			(2.466)			(2.384)
HI			6.093**			1.652
			(2.320)			(1.762)
VI			3.205			3.928
			(3.315)			(2.336)
OI			3.139			-1.244
			(2.201)			(2.023)
OHI			4.679*			0.700
			(2.069)			(2.133)
OLI			-1.708			-3.196
			(2.347)			(2.453)
Black			-0.851			-1.284
			(1.732)			(1.932)
Hispanic			2.812			2.546
			(2.080)			(1.801)
Asian			1.343			2.220
			(1.848)			(1.522)
Other			-4.520			-1.422
			(4.507)			(6.816)
Male			1.799			2.437
			(1.006)			(1.264)
EL			-1.502			-0.216
			(1.972)			(1.637)

Table 5, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
Suspension			-1.009			-1.214
			(1.190)			(1.218)
FRPL			-0.140			-1.291
			(1.054)			(1.079)
Rural			4.386*			4.870**
			(1.901)			(1.500)
Suburban			1.121			1.253
			(1.405)			(1.297)
_cons	135.641***	135.778***	139.151***	189.433***	188.894***	201.295***
	(9.012)	(8.923)	(10.112)	(19.511)	(19.646)	(19.844)
Random Effect						
Student	29.126***	28.663***	17.494***	49.572***	49.583***	32.330***
	(7.907)	(8.082)	(6.045)	(14.320)	(14.335)	(9.076)
Residual	175.842***	175.818***	164.404***	118.042***	117.920***	117.251***
	(16.923)	(16.805)	(17.813)	(10.583)	(10.533)	(10.980)
N	2759	2759	2477	2623	2623	2359
ll	-7891885	-7889956	-6882410	-7313014	-7312300	-6384193

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5, cont.

First Grade

Letter word identification. In table 6, results also showed a significant initial linear slope in the trajectories of letter word identification scores for the promoted group at first grade, γ_{10} = 7.39, p<.001, which was modified by a significant quadratic negative effect, γ_{20} = -0.50, p< .001. Student age at the initial Wave is a significant negative predictor (γ_{09} = -8.99, p< .001) and baseline performance (γ_{10} = 0.91, p< .001) is a significant positive predictor in the first model. The significant effects of student age at the beginning of the grade and baseline measurement persist across all three models. The effect of retention at first grade was negative but nonsignificant across all three models, indicating no significant difference on letter word identification between students retained at first grade versus those who were promoted. In addition, none of the interactions between retention status at first grade and age reaches significance level, indicating the achievement gaps between the two groups do not change over time. Instead of retention status at first grade, suspension history produced significant effects for the growth curves of letter word identification performance among students with disabilities. For instance, students who were suspended before scored on average 3.65 points significantly lower than their counterparts while holding all the other predictors constant.

Passage comprehension. Results in table 6 indicate the average level was significant and the average passage comprehension scores for the reference group (white non-EL male students, not enrolled in free or reduced-price lunch programs, have autism, never suspended before, live in urban areas and promoted to next grade level) were 164.15 at age 12.22 while adjusting for baseline passage comprehension performance and student age at initial Wave. The acceleration rate of yearly gain of passage comprehension also slightly decreased in each subsequent year as the significant initial linear slope γ_{10} = 3.99,

p<.001 is modified by the significant negative quadratic effect γ_{20} = -0.35, p<.001. Students retained at first grade scored no significantly different than their promoted peers and this is true for all three models. None of the interactions between retention status and age is significant, indicating the gaps on passage comprehension between the two groups do not change over time. Model 3 indicates disability type and suspension history are significant predictors while holding other covariates constant. For instance, students with SLD, SLI, EBD, VI, OI and OHI scored 5.33, 6.76, 7.30, 5.86, 5.03, and 4.76 points significantly higher on the outcome than their peers with autism. In contrast, students with disabilities who were suspended before scored on average 2.28 points significantly lower than their peers who were never suspended before.

Applied problem. Results in table 7 indicate the average level was significant and the applied problem scores for the reference group (white non-EL male students, not enrolled in free or reduced-price lunch programs, have autism, never suspended before, live in urban areas and promoted to next grade level) were 133.77 at age 12.22 while adjusting for baseline applied problem performance and student age at the initial Wave. The significant effects of baseline measurement and initial age persist across all three models. The effect of retention at first grade is not significant even after controlling for the effects of other covariates, indicating retained students scored at no statistically different level as opposed to their promoted peers. In addition, none of the interactions between retention status and age or age squared is significant, indicating the gaps on applied problem between the retained and promoted groups at first grade remain constant over time. Similar to the significant predictors identified at kindergarten, disability type and family location produced significant coefficients. For instance, students with SLD, SLI, EBD, HI and OHI scored 3.83, 6.47, 6.39, 5.91 and 4.68 points significantly higher than

their peers with autism. Compared to students with disabilities who lived in the urban area, students who lived in the rural areas scored on average 4.17 points significantly higher.

Calculation. Results in table 7 indicate that the average level was significant and the calculation scores for the reference group (white non-EL male students, not enrolled in free or reduced-price lunch programs, have autism, never suspended before, live in urban areas and promoted to next grade level) were 188.94 at age 12.22 while adjusting for baseline calculation performance and student age at the initial Wave. Similarly, the significant effects of baseline performance and initial age persist across all three models. Students retained at first grade scored no significantly different than their promoted peers across three models, indicating retained students performed at similar levels on calculation than their promoted peers at first grade. In addition, none of the interaction terms between retention status at first grade and age is significant, indicating the trajectories of calculation between the two groups do not change over time. Significant predictors of the outcome include disability type and family location. For instance, students with SLI scored on average 5.94 points significantly higher on calculation than students with autism. Students with disabilities living in the rural areas scored on average 4.66 points significantly higher on calculation than their peers living in the urban areas.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
Age	7.387***	7.647***	7.214***	3.994***	3.968***	3.967***
	(0.341)	(0.369)	(0.367)	(0.260)	(0.303)	(0.293)
Age^2	-0.502***	-0.438***	-0.430***	-0.350***	-0.336***	-0.357***
	(0.068)	(0.083)	(0.077)	(0.071)	(0.101)	(0.070)
Age at initial wave	-8.986***	-9.008***	-8.320***	-4.715***	-4.715***	-4.493***
	(0.542)	(0.545)	(0.522)	(0.368)	(0.368)	(0.397)
LW at initial wave	0.913***	0.914***	0.895***	-	-	-
	(0.020)	(0.020)	(0.020)			
PC at initial wave	-	-	-	0.779***	0.778***	0.778***
				(0.024)	(0.024)	(0.029)
Retained at 1st	-1.201	-0.255	-2.015	-0.972	-0.813	-1.195
	(1.136)	(1.284)	(1.227)	(0.886)	(1.273)	(0.922)
Retained at 1st # age		-0.771			0.055	
		(0.494)			(0.380)	
Retained at 1st # age^2		-0.196			-0.030	
		(0.132)			(0.127)	
SLD			0.125			5.327***
			(2.187)			(1.496)
SLI			0.468			6.758***
			(2.351)			(1.641)

Table 6. Hierarchical Linear Regression Model Predicting Reading Performance (Retained at First Grade)

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
MR			-0.737			3.328
			(2.426)			(1.699)
EBD			2.330			7.300***
			(2.405)			(2.046)
HI			0.640			2.412
			(2.237)			(1.513)
VI			1.409			5.864**
			(2.549)			(1.938)
OI			-0.336			5.031***
			(2.485)			(1.519)
OHI			2.979			4.759**
			(2.431)			(1.558)
OLI			-2.066			1.078
			(2.103)			(1.595)
Black			-1.091			-1.525
			(1.266)			(1.400)
Hispanic			4.068			-1.237
			(2.730)			(1.988)
Asian			-0.151			1.895
			(2.221)			(1.955)
Other			-2.682			-9.933
			(2.297)			(5.318)
Male			1.712			0.549
			(1.242)			(0.962)

Table 6, cont.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
EL			-2.611			1.577
			(2.988)			(1.817)
Suspension			-3.654**			-2.275*
			(1.118)			(0.997)
FRPL			-0.754			-0.431
			(1.211)			(1.182)
Rural			1.287			1.302
			(1.409)			(1.160)
Suburban			-0.516			1.481
			(1.200)			(1.071)
_cons	150.380***	149.667***	151.769***	164.158***	164.249***	156.895***
	(9.585)	(9.551)	(10.756)	(10.881)	(10.773)	(12.747)
Random Effect						
Student	45.904***	47.175***	30.744***	23.881***	23.932***	15.395***
	(11.710)	(12.234)	(6.122)	(7.250)	(7.260)	(5.773)
Residual	151.838***	150.060***	148.882***	117.584***	117.533***	117.264***
	(14.507)	(14.212)	(15.890)	(13.120)	(13.128)	(14.438)
N	2773	2773	2491	2847	2847	2552
ll	-7838410	-7833650	-6853729	-7565271	-7565186	-6633990

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
Age	4.632***	4.853***	4.658***	4.316***	4.611***	4.397***
	(0.358)	(0.365)	(0.355)	(0.359)	(0.342)	(0.389)
Age^2	-0.298***	-0.248*	-0.348***	-0.463***	-0.499***	-0.480***
	(0.071)	(0.105)	(0.070)	(0.068)	(0.093)	(0.076)
Age at initial wave	-4.304***	-4.312***	-4.478***	-4.272***	-4.289***	-3.831***
	(0.509)	(0.500)	(0.491)	(0.533)	(0.522)	(0.616)
AP at initial wave	0.835***	0.836***	0.821***	-	-	-
	(0.020)	(0.020)	(0.022)			
CAL at initial wave	-	-	-	0.730***	0.732***	0.693***
				(0.042)	(0.041)	(0.044)
Retained at 1st	0.959	1.716	0.835	-1.713	-2.004	-1.804
	(1.191)	(1.420)	(1.166)	(1.347)	(1.692)	(1.370)
Retained at 1st # age		-0.658			-0.760	
		(0.470)			(0.475)	
Retained at 1st # age^2		-0.158			0.042	
		(0.144)			(0.133)	
SLD			3.826*			0.592
			(1.948)			(1.802)
SLI			6.471**			5.942**
			(2.108)			(2.001)

Table 7. Hierarchical Linear Regression Model Predicting Math Performance (Retained at First Grade)

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
MR			-0.337			-2.316
			(2.041)			(2.266)
EBD			6.392**			2.097
			(2.463)			(2.440)
HI			5.909*			1.597
			(2.326)			(1.722)
VI			3.003			3.728
			(3.301)			(2.431)
OI			2.792			-1.695
			(2.195)			(1.997)
OHI			4.678*			0.603
			(2.049)			(2.116)
OLI			-1.755			-4.022
			(2.344)			(2.318)
Black			-0.546			-0.600
			(1.708)			(1.968)
Hispanic			2.826			3.553
			(2.068)			(1.861)
Asian			1.871			2.128
			(1.975)			(1.618)
Other			-4.933			-2.665
			(4.455)			(6.834)
Male			1.764			2.391
			(1.017)			(1.327)

Table 7, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
EL			-1.141			-0.072
			(1.934)			(1.599)
Suspension			-0.941			-1.654
			(1.210)			(1.370)
FRPL			-0.125			-1.521
			(1.063)			(1.126)
Rural			4.171*			4.664**
			(1.882)			(1.516)
Suburban			1.010			1.178
			(1.383)			(1.262)
_cons	133.768***	133.390***	136.789***	188.940***	188.049***	199.319***
	(9.057)	(8.964)	(9.434)	(19.477)	(19.344)	(19.743)
Random Effect						
Student	28.969***	28.142***	17.330***	49.388***	49.875***	32.212***
	(7.904)	(7.752)	(6.012)	(14.375)	(14.417)	(9.362)
Residual	176.262***	176.280***	164.651***	118.184***	117.292***	117.764***
	(16.950)	(16.807)	(17.891)	(10.692)	(10.269)	(11.198)
N	2759	2759	2477	2623	2623	2359
ll	-7893313	-7890153	-6882915	-7313179	-7309522	-6386695

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 7, cont.

Second Grade

Letter word identification. In table 8, the average level was significant and indicated that the average letter word identification scores for the reference group (white non-EL male students, not enrolled in free or reduced-price lunch programs, have autism, never suspended before, live in urban areas and promoted to next grade level) were 149.71 at age 12.22 while adjusting for baseline letter word identification and student age at the initial Wave. Results also showed a significant linear effect γ_{10} = 7.41, p<.001 modified by a negative quadratic effect, γ_{20} = -0.51, p<.001. Students retained at second grade scored no significantly different than their promoted peers in the base model when retention status and age are the only two predictors. Once controlling for the effect of all the other covariates, retained group still scored at no statistical difference compared to their promoted peers given the coefficient associated with retention remains nonsignificant. Besides baseline measurement and initial student age, suspension history is also a significant predictor in the full model. Students with disabilities who were suspended before scored on average 3.23 points significantly lower than their peers who were never suspended.

Passage comprehension. Table 8 shows the linear change in passage comprehension was positive and significant, γ_{10} = 4.01, p< .001, indicating that students with disabilities who were promoted grew 4.01 score points on passage comprehension at age 12.22 than students retained at second grade. Similarly, baseline measurement and student age at the initial Wave are significant predictors across all three models. Students who were one year older at the beginning of grade scored 4.72 points significantly lower whereas students who scored one unit higher at the baseline on passage comprehension scored 0.78 points significantly higher. Students retained at second grade scored no significant different than their promoted peers. In addition, none of the interactions

between retention and age or age squared reaches significant level, indicating the gaps on passage comprehension trajectories between the two groups do not differ significantly over time. Disability types and suspension history are significant predictors of the development of passage comprehension. For instance, compared to students with autism, students with SLD, SLI, EBD, VI, OI and OHI scored 5.22, 6.86, 7.13, 5.89, 5.04 and 4.85 points significantly higher on passage comprehension. Students who were suspended before scored 2.02 points significantly lower than their peers who were never suspended.

Applied problem. Table 9 shows the linear change in applied problem was positive and significant, γ_{10} = 4.62, p< .001, indicating that students with disabilities who were promoted grew 4.62 score points on applied problem at age 12.22 than students retained at second grade while adjusting for baseline measurement and student age at the initial Wave. Students retained at second grade scored no significantly different than their promoted peers across three models. The interactions between retention status at second grade and age are nonsignificant, indicating the trajectories of applied problems between the two groups do not change over time. Student disability types and family location are significant predictors of applied problem performances for students with disabilities. Compared to students with autism, students with SLD, SLI, EBD, HI and OHI scored 3.95, 6.56, 6.46, 5.98 and 4.67 points significantly higher on the outcome while holding all the other predictors constant. In addition to disability types, students with disabilities living in the rural areas scored on average 4.23 points significantly higher on applied problem compared to their peers living in the urban areas.

Calculation. Results in table 9 shows the linear change in applied problem was positive and significant, γ_{10} = 4.34, p< .001, indicating that students with disabilities who were promoted grew 4.34 score points on calculation at age 12.22 than students retained at second grade while adjusting for baseline calculation performance and student age at the

initial Wave. The significant effects of baseline measurement and initial student age persist across all three models. Students retained at second grade scored no significant different than their promoted peers across all three models. None of the interactions between retention status and student age is significant, indicating the trajectories of calculation performance between the two groups do not change over time. Compared to students with autism, students with SLI scored 6.41 points significantly higher on the outcome than students with autism. In addition, students with disabilities living in the rural areas scored on average 4.56 point significantly higher than their peers living in the urban areas.

Academic Trajectories Using Full Model Results

The upper left panel in figure 1 through figure 3 displays the trajectories of letter word identification by retention status at kindergarten, first and second grade from student age 7 to 17 years old using the base model result. Students from both groups showed very similar slopes and accelerations in letter word identification growth, and the rate of growth plateaued with age. The group difference in letter word identification by retention status at kindergarten is negligible, as reflected in the nonsignificant coefficient associated with retention from model results.

The upper right plot in figure 1 through 3 displays the trajectories of passage comprehension by retention status at kindergarten, first and second grade from age 7 to 17 years old using the full model result. Students retained at kindergarten, first and second grade are performing at a very similar level as opposed to those who were promoted and the achievement trajectory gap between the two groups stays constant, indicating there is no significant interactions between retention status and student age, reflected in model 2. The slope and rate of trajectories for passage comprehension are similar to that of letter word identification in that both trajectories become plateaued as students grow older.

The bottom left plot in figure 1 through figure 3 shows the trajectories of applied problem between students retained at kindergarten, first and second grade versus their promoted peers. Retained students show similar performance trajectories to their promoted peers on applied problem at kindergarten, first and second grade. In addition, the gaps between the two groups stay constant in three plots, reflected the nonsignificant interactions between retention status and student age.

The bottom right plots in figure 1 through figure 3 shows the trajectories of calculation between students retained at kindergarten, first and second grade as opposed to their promoted peers. Students retained at kindergarten had much lower trajectories on calculation compared to their promoted peers. For students retained at first and second grade, although they consistently performed at a lower level as opposed to their promoted peers, the group differences are much smaller, almost neglectable. In addition, the gaps between the two groups stay constant, reflected the nonsignificant interactions between retention status and student age.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
Age	7.405***	7.321***	7.228***	4.009***	4.041***	3.975***
	(0.338)	(0.417)	(0.367)	(0.262)	(0.286)	(0.294)
Age^2	-0.506***	-0.489***	-0.439***	-0.353***	-0.371***	-0.362***
	(0.069)	(0.084)	(0.078)	(0.070)	(0.072)	(0.071)
Age at initial wave	-8.995***	-8.982***	-8.298***	-4.718***	-4.729***	-4.467***
	(0.538)	(0.538)	(0.536)	(0.359)	(0.360)	(0.396)
LW at initial wave	0.914***	0.915***	0.896***			
	(0.019)	(0.020)	(0.020)			
PC at initial wave				0.779***	0.779***	0.779***
				(0.025)	(0.024)	(0.029)
Retained at 2nd	-0.503	-0.250	-1.050	-0.537	-0.820	-0.312
	(1.616)	(1.642)	(1.226)	(1.047)	(1.253)	(0.922)
Retained at 2nd # age		0.206			-0.065	
		(0.775)			(0.486)	
Retained at 2nd # age^2		-0.047			0.055	
		(0.170)			(0.147)	
SLD			-0.036			5.220***
			(2.093)			(1.464)
SLI			0.733			6.858***
			(2.302)			(1.628)

Table 8. Hierarchical Linear Regression Model Predicting Reading Performance (Retained at Second Grade)

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
MR			-0.773			3.276
			(2.386)			(1.673)
EBD			2.050			7.126***
			(2.333)			(1.995)
HI			0.462			2.294
			(2.154)			(1.472)
VI			1.400			5.887**
			(2.411)			(1.865)
OI			-0.206			5.043***
			(2.369)			(1.486)
OHI			3.201			4.853**
			(2.376)			(1.509)
OLI			-1.447			1.362
			(1.998)			(1.592)
Black			-1.117			-1.577
			(1.281)			(1.381)
Hispanic			3.349			-1.618
			(2.805)			(2.022)
Asian			0.122			2.185
			(2.120)			(2.007)
Other			-2.837			-9.883
			(2.471)			(5.276)
Male			1.638			0.552
			(1.257)			(0.938)

Table 8, cont.

	Letter word identification			Passage comprehension		
	M1	M2	M3	M4	M5	M6
EL			-2.142			1.810
			(2.958)			(1.836)
Suspension			-3.232**			-2.022*
			(1.121)			(1.001)
FRPL			-0.625			-0.346
			(1.219)			(1.185)
Rural			1.217			1.241
			(1.407)			(1.171)
Suburban			-0.559			1.462
			(1.187)			(1.077)
_cons	149.708***	149.358***	150.585***	163.819***	164.058***	155.840***
	(9.812)	(9.886)	(10.910)	(11.009)	(10.987)	(12.697)
Random Effect						
Student	46.293***	45.731***	31.838***	24.236***	24.230***	16.077***
	(11.818)	(11.721)	(5.836)	(7.160)	(7.119)	(5.819)
Residual	151.754***	152.083***	148.528***	117.424***	117.399***	116.930***
	(14.485)	(14.543)	(15.861)	(13.072)	(13.070)	(14.372)
N	2773	2773	2491	2847	2847	2552
ll	-7839393	-7839066	-6856082	-7566079	-7565865	-6635515

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
Age	4.621***	4.479***	4.652***	4.337***	4.352***	4.405***
	(0.355)	(0.413)	(0.354)	(0.358)	(0.419)	(0.388)
Age^2	-0.298***	-0.269**	-0.348***	-0.468***	-0.401***	-0.493***
	(0.071)	(0.086)	(0.070)	(0.068)	(0.078)	(0.077)
Age at initial wave	-4.334***	-4.318***	-4.506***	-4.257***	-4.232***	-3.774***
	(0.511)	(0.506)	(0.495)	(0.544)	(0.547)	(0.650)
AP at initial wave	0.834***	0.835***	0.819***	-	-	-
	(0.020)	(0.020)	(0.022)			
CAL at initial wave	-	-	-	0.730***	0.733***	0.688***
				(0.043)	(0.042)	(0.046)
Retained at 2nd	-0.577	-0.132	-0.492	-0.247	0.826	-1.899
	(0.970)	(1.096)	(1.075)	(1.376)	(1.473)	(1.030)
Retained at 2nd # age		0.360			-0.125	
		(0.497)			(0.527)	
Retained at 2nd # age^2		-0.083			-0.218	
		(0.141)			(0.138)	
SLD			3.954*			0.438
			(1.921)			(1.834)
SLI			6.564**			6.405**
			(2.109)			(2.046)

Table 9. Hierarchical Linear Regression Model Predicting Math Performance (Retained at Second Grade)

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
MR			-0.274			-2.432
			(2.031)			(2.301)
EBD			6.462**			1.699
			(2.449)			(2.401)
HI			5.981**			1.334
			(2.302)			(1.735)
VI			3.048			3.783
			(3.276)			(2.334)
OI			2.953			-1.380
			(2.193)			(1.997)
OHI			4.666*			0.817
			(2.050)			(2.099)
OLI			-1.841			-3.449
			(2.332)			(2.414)
Black			-0.564			-0.727
			(1.720)			(1.958)
Hispanic			2.885			2.517
			(2.120)			(1.777)
Asian			1.397			2.069
			(1.990)			(1.736)
Other			-5.298			-3.439
			(4.528)			(6.900)
Male			1.707			2.186
			(1.038)			(1.276)

Table 9, cont.

	Applied problem			Calculation		
	M1	M2	M3	M4	M5	M6
EL			-1.230			0.446
			(1.964)			(1.552)
Suspension			-1.081			-1.298
			(1.191)			(1.278)
FRPL			-0.124			-1.237
			(1.041)			(1.076)
Rural			4.234*			4.586**
			(1.883)			(1.533)
Suburban			1.060			1.151
			(1.404)			(1.285)
_cons	135.194***	134.381***	138.339***	187.991***	186.083***	201.175***
	(9.129)	(9.116)	(9.837)	(19.470)	(19.218)	(20.365)
Random Effect						
Student	29.110***	29.095***	17.385***	50.526***	49.892***	33.265***
	(7.819)	(7.835)	(5.893)	(14.280)	(14.327)	(9.239)
Residual	176.256***	176.070***	164.686***	117.877***	117.893***	116.966***
	(16.960)	(16.792)	(17.883)	(10.605)	(10.796)	(10.927)
N	2759	2759	2477	2623	2623	2359
ll	-7893832	-7892871	-6883311	-7315685	-7313329	-6386365

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 9, cont.

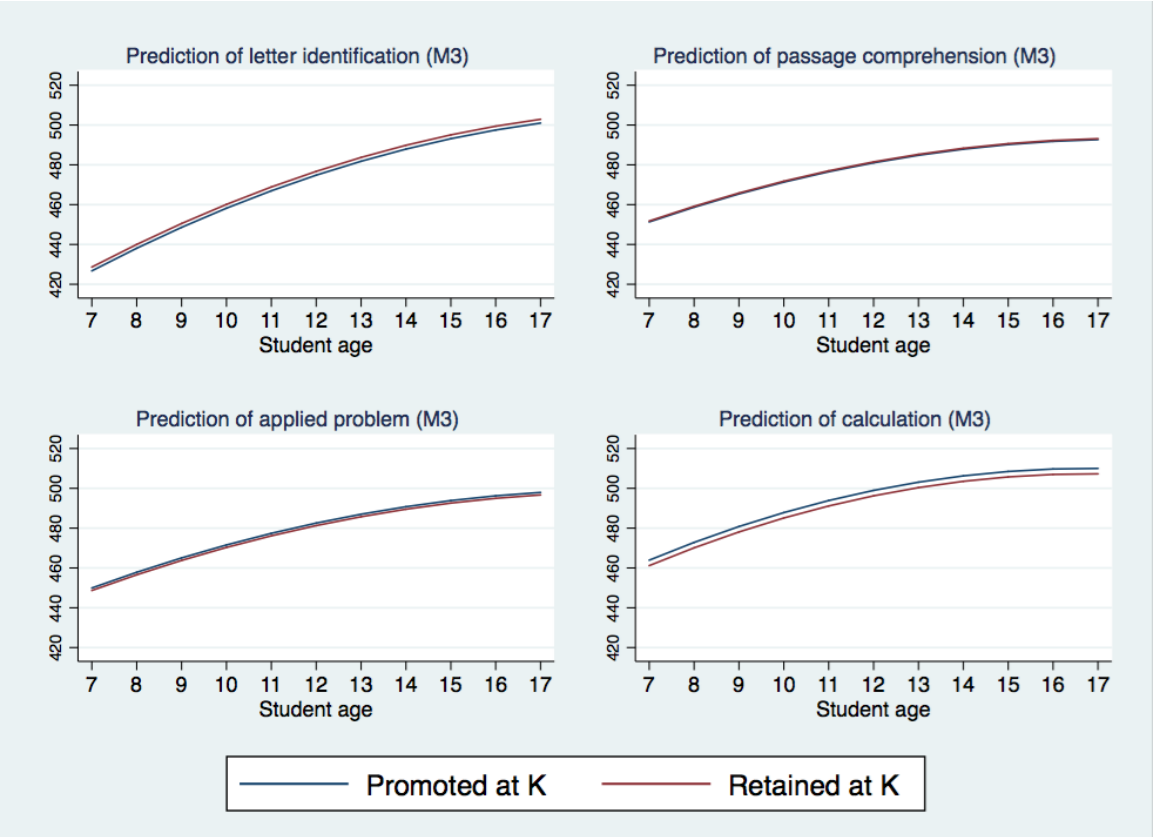

Figure 1: Predicted Reading and Math Performance by Retention Status at Kindergarten Using Full Model Results

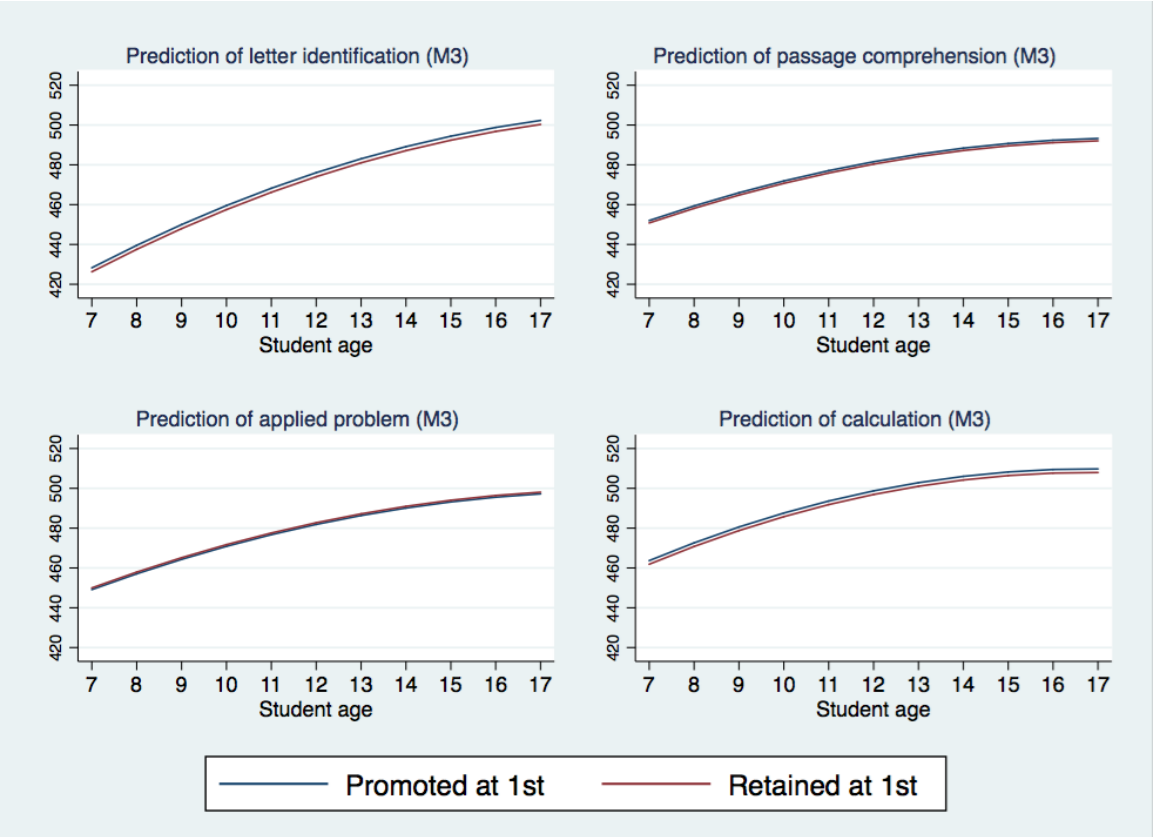

Figure 2: Predicted Reading and Math Performance by Retention Status at First Grade Using Full Model Results

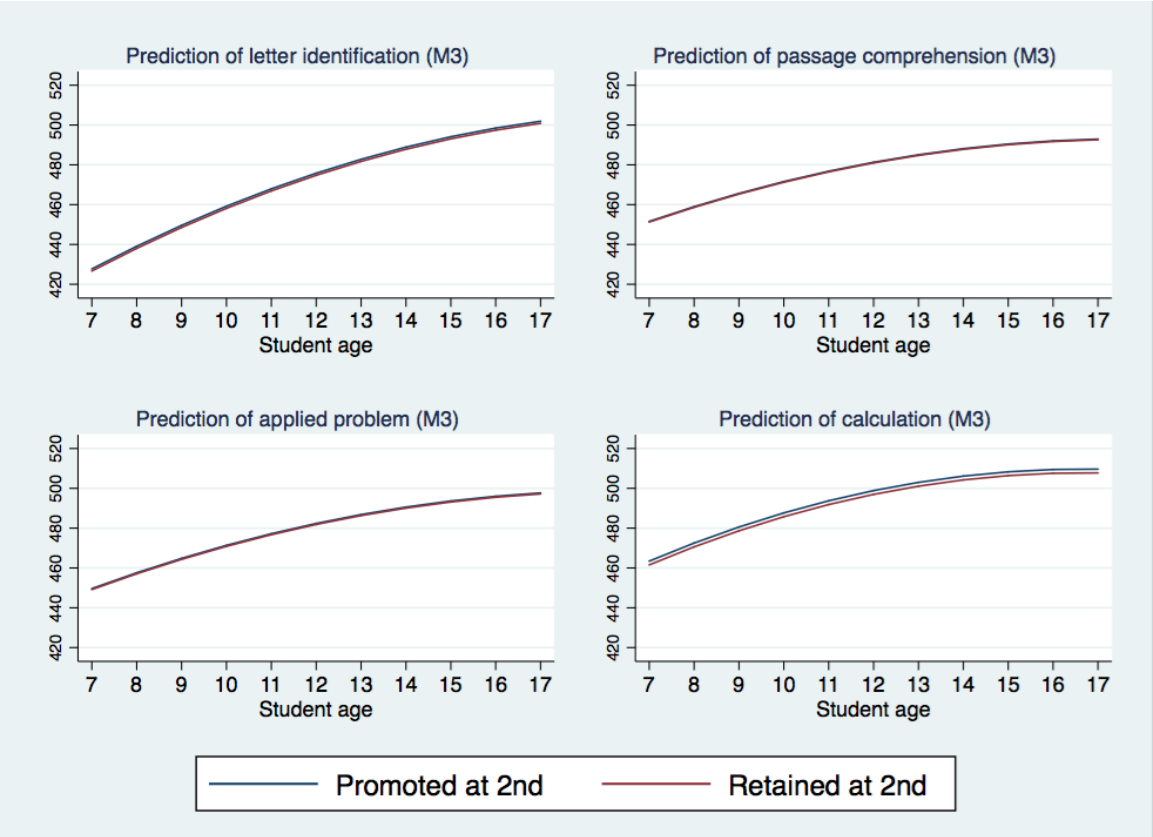

Figure 3: Predicted Reading and Math Performance by Retention Status at Second Grade Using Full Model Results

SOCIAL-BEHAVIORAL TRAJECTORIES OF STUDENTS EVER RETAINED AT KINDERGARTEN, FIRST AND SECOND GRADE

Table 10 presents the hierarchical linear regression model for estimating growth curves in the composite social-behavioral measures for students with disabilities from ages 7 to 17. For retention status at each of the three grade levels, two models were computed to predict social-behavioral outcome. Model 1 contains age, age squared, student age at the initial Wave and retention status at each grade level, and model 2 adds all the other covariates such as disability types, student race, gender, EL status, suspension history, free or reduced-priced lunch enrollment and home geography. Figure 4 presents the estimated predictions of social-behavioral predictions by retention status using two model results for each grade level.

Kindergarten

Results in table 10 show a one-year change in student age results in a significant 0.03 lower standardized social-behavioral score for promote students at age 12.22 while adjusting for baseline student age. In addition, the relationship between student age and the social-behavioral outcome is linear as the quadratic age is not significant. Retained students at kindergarten scored similarly on social-behavioral outcome than their promoted peers as the coefficient associated with retention is not significant. Once controlling for all the other covariates, the coefficient associated with retention at kindergarten is getting smaller but still nonsignificant and the change of social-behavioral outcome depends more on student disability types, gender, ELL status and suspension history. Compared to students with autism, students with SLD, SLI, MR, HI, OI and OHI produced significant positive coefficients, indicating better social-behavioral performance. Boys scored on average 0.13 points lower than girls on social-behavioral outcome and the difference is significant. ELs scored about 0.22 points significant lower than their native speaker peers. Students

suspended before scored about 0.22 points lower than their peers who were never retained before while holding all the other covariates constant.

First Grade

Similar to what I found of retention at kindergarten, a significant linear effect was found (γ_{10} = -0.03, p<.05) at first grade, indicating the trajectory of student social-behavioral outcome is linear. Students retained at first grade scored no significant different than their promoted peers in terms of social-behavioral outcome. Once controlling for all other predictors, there is still no statistically significant relationship between retention status and the outcome. Findings regarding the significant predictors of student social-behavioral outcome while controlling for the effect of retention at first grade and student age remain consistent as that of kindergarten. While holding all the other predictors constant, students with SLD, SLI, MR, HI, OI, and OHI scored about 0.32, 0.40, 0.27, 0.37, 0.25 and 0.18 points significant higher than their peers with autism. Compared to girls, boys scored 0.13 points significant lower, ELs scored about 0.23 points significant lower than their native speaker peers and those who were suspended scored 0.23 points significant lower than their peers who were never suspended.

Second Grade

Results of second grade indicate that both the linear and quadratic student age produced nonsignificant coefficient, indicating the change of social-behavioral outcome is not statistically related to the change of student age. Students retained at second grade scored no significant different as opposed to their promoted peers and this is true even after controlling for all the other predictors. Compared to students with autism, students with SLD, SLI, MR, HI, OI and OHI scored 0.31, 0.40, 0.26, 0.37, 0.24 and 0.18 points significantly higher. Besides, gender, EL and suspension history also produced significant

results. Boys, ELs and students who were ever suspended scored on average 0.12, 0.23 and 0.22 points significantly lower than girls, native speaker students and students who were never suspended while holding all the other predictors constant.

Figure 4 displays the prediction of social-behavioral outcome by retention status at kindergarten, first and second grade using both model results. Plots of social-behavioral trajectories at kindergarten show that student's social-behavioral outcome is getting worse across student ages from 7 to 17 years old but retained students at kindergarten produced greater social-behavioral outcome than their promoted peers. However, the gaps between retained and promoted groups at first and second grade are much smaller and almost negligible. Once controlling for other predictors, the plots show that the gaps between retained and promoted groups at kindergarten, first and second grade all get larger, indicating that the group gaps widen up once taking into account other predictors. However, retained students produced worse social-behavioral outcome than promoted peers at kindergarten once controlling for other predictors whereas retained students at first and second grade produced better outcome than their promoted peers.

	Kindergarten		1st grade		2nd grade	
	M1	M2	M3	M4	M5	M6
Age	-0.025	-0.023	-0.025	-0.023	-0.024	-0.023
	(0.013)	(0.014)	(0.013)	(0.014)	(0.013)	(0.014)
Age^2	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age at initial wave	0.009	0.017	0.010	0.014	0.011	0.017
	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
Retained	0.019	0.004	0.005	-0.070	0.074	0.047
	(0.060)	(0.059)	(0.051)	(0.048)	(0.054)	(0.055)
SLD		0.308^{***}		0.316^{***}		0.309^{***}
		(0.086)		(0.087)		(0.087)
SLI		0.406^{***}		0.402^{***}		0.400^{***}
		(0.095)		(0.095)		(0.095)
MR		0.261^{**}		0.265^{**}		0.260^{**}
		(0.092)		(0.091)		(0.092)
EBD		0.119		0.128		0.125
		(0.098)		(0.098)		(0.098)
HI		0.367^{***}		0.373^{***}		0.371^{***}
		(0.090)		(0.091)		(0.092)
VI		0.227		0.225		0.228
		(0.124)		(0.123)		(0.124)
OI		0.252^{**}		0.253^{**}		0.243^{*}
		(0.096)		(0.097)		(0.098)

Table 10. Hierarchical Linear Regression Model Predicting Social-behavioral Outcome Using Retention at Kindergarten, First and Second Grade as Predictors

	Kindergarten		1st grade		2nd grade	
	M1	M2	M3	M4	M5	M6
OHI		0.185*		0.181*		0.184*
		(0.083)		(0.084)		(0.084)
OLI		0.120		0.107		0.115
		(0.089)		(0.090)		(0.089)
Black		-0.087		-0.088		-0.088
		(0.089)		(0.086)		(0.087)
Hispanic		0.130		0.146		0.143
		(0.111)		(0.109)		(0.110)
Asian		-0.046		-0.066		-0.034
		(0.081)		(0.083)		(0.083)
Other		-0.067		-0.076		-0.045
		(0.125)		(0.119)		(0.124)
Male		-0.127*		-0.130*		-0.122*
		(0.054)		(0.054)		(0.055)
EL		-0.222*		-0.234*		-0.230*
		(0.112)		(0.112)		(0.112)
Suspension		-0.221***		-0.233***		-0.223***
		(0.057)		(0.056)		(0.057)
FRPL		-0.089		-0.091		-0.094
		(0.054)		(0.053)		(0.054)
Rural		0.007		0.011		0.006
		(0.077)		(0.078)		(0.078)
Suburban		0.070		0.072		0.068
		(0.062)		(0.062)		(0.063)

Table 10, cont.

	Kindergarten		1st grade		2nd grade	
	M1	M2	M3	M4	M5	M6
_cons	-0.044	-0.206	-0.043	-0.149	-0.087	-0.231
	(0.206)	(0.214)	(0.211)	(0.221)	(0.206)	(0.215)
Random Effect						
Student	0.128***	0.092***	0.128***	0.091***	0.128***	0.092***
	(0.016)	(0.016)	(0.016)	(0.015)	(0.016)	(0.015)
Residual	0.174***	0.173***	0.174***	0.173***	0.174***	0.172***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
N	2255	2149	2255	2149	2255	2149
ll	-1142619	-1024832	-1142726	-1022768	-1140651	-1023903

Note. Reference group = Promoted, Autism, White, Female, non-EL, never suspended, not enrolled in free or reduced-price lunch program, live in the urban area. OLI= other low incidence disabilities. Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10, cont.

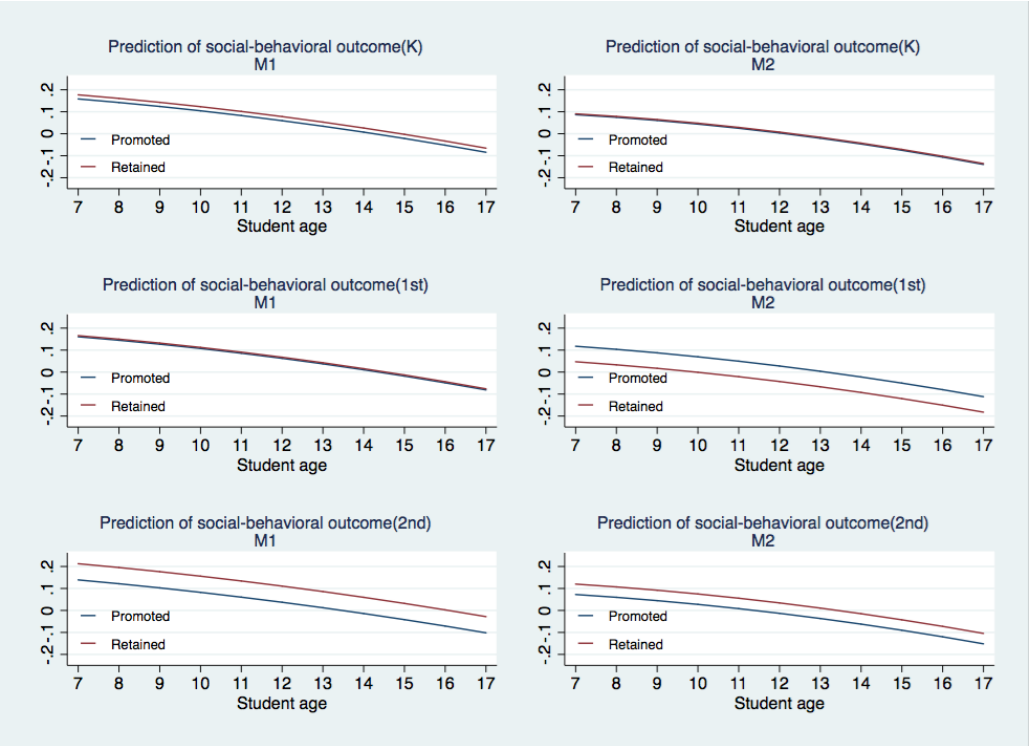

Figure 4: Predicted Composite Social-behavioral Outcome by Retention Status at Kindergarten, First and Second Grade

Table 11 through 15 present standardized differences for a total of 15 predictors included in the propensity score matching. Standardized difference was obtained by dividing the difference between the mean of a baseline variable in each group by the squared root of the sum of variances over 2. A noteworthy point is that in small matched samples (i.e., around or less than 100), which is the case of this study, moderate standardized differences can be expected when propensity score model is correctly specified (Austin, 2009). Therefore, I used a moderate (-0.234 to 0.234) standardized difference cut-off when checking the balance of covariates between groups after matching.

Propensity Matching Results on Retention Status From Second to Fifth Grade

A nearest-neighbor matching with replacement using previously identified four groups of student subsamples was conducted at each grade level from Grade 2 through 5. Before matching on retention at second grade, a total of 139 students were identified for second grade group. I constructed a total of 108 matched pairs using 73 promoted and 66 retained students. The absolute standardized differences for the covariates after matching range from 0 to 0.177 with variance ratios all close to 1. The top left plot in figure 5 presents density plots for retained and promoted students from both the raw (unmatched) and matched cases and the two groups were more closely equated on their propensity scores following the matching process as opposed to the raw cases in that the two distributions in matched cases almost overlapping each other.

A total of 107 students were identified for third grade group before the matching. I constructed a total of 84 matched sets using 69 promoted and 38 retained students. The absolute standardized difference for the covariates range from 0 to 0.236 with all variance ratios close to 1. In addition, 14 out of 15 predictors were well matched except for student

self-care skills at initial Wave which produced a slightly greater standardized difference (0.236) as opposed to the cut-off. The top right plot in figure 5 presents density plots for the two groups before and after matching. The distribution of propensity scores in matched cases between the two groups are overlapping each other way better than that in unmatched cases.

A total of 105 students were identified for fourth grade group before the matching and I constructed a total of 83 matched sets using 70 promoted and 35 retained students. The absolute standardized difference for the covariates after matching range from 0.021 to 0.268 and 14 out of 15 predictors were well matched except for student self-care skills at the initial Wave which produced a greater value of 0.268 as opposed to the cut-off. The bottom left plot in figure 5 displays density plots for the two groups in unmatched and matched cases. The distribution of propensity scores in matched cases between the two groups were overlapped better.

A total of 133 students were identified for the fifth-grade group before the matching and I constructed a total of 110 matched sets using 90 promoted and 43 retained students. The absolute standardized difference for all covariates after matching range from 0.011 to 0.246 and all covariates except for EL status, which produced a slightly greater standardized difference as opposed to the cut-off, were well matched. The bottom right plot in Figure 5 displays the density plots for the two groups before and after matching and the distribution of propensity scores in the matched cases produced better overlapping between the two groups.

	Second grade			
	Before matching (N=139)		After matching (N=139)	
Child demographics	Standardized differences	Variance ratio	Standardized differences	Variance ratio
Gender	0.013	0.994	0.000	1.000
Disability	-0.059	0.913	-0.173	0.734
Student race	-0.040	1.115	-0.156	0.957
EL	-0.114	0.806	-0.040	0.925
Suspension history	-0.025	0.961	-0.109	1.185
Family income	0.197	1.136	0.110	1.007
Overall health	-0.118	1.030	0.087	0.971
Maternal education	0.084	1.153	0.056	1.101
Participated in at least one extracurricular activity	0.281	0.859	-0.015	1.009
Frequency of reading together per week	0.047	0.902	-0.041	1.087
Baseline measure				
Functional skills	0.150	0.915	0.158	0.813
Social skills	0.052	0.735	-0.093	0.703
Living skills	-0.295	0.825	-0.024	1.068
Self-care skills	-0.255	1.203	-0.089	1.091
Mental skills	-0.160	0.864	-0.177	0.757

Table 11. Standardized Mean Difference and Variance Ratio for Predictors Before and After Matching on Retention at Second Grade

Note. Functional, social, living, self-care and mental skills are baseline measures extracted from Wave 1. The standardized difference for covariate z between treatment (t) and control groups (t_0) are computed as $\delta_z(t) = [\widehat{\mu_z}(t) - \widehat{\mu_z}(t_0)] / \sqrt{\frac{\widehat{\sigma}_z(t)^2 + \widehat{\sigma}_z(t_0)^2}{2}}$. The variance ratio is computed as $\rho_z(t) = \{\hat{\sigma}_z^2(t)\}/\{\hat{\sigma}_z^2(t_0)\}$.

	Third grade			
	Before matching (N=107)		After matching (N=107)	
Child demographics	Standardized differences	Variance ratio	Standardized differences	Variance ratio
Gender	-0.188	1.142	0.100	0.922
Disability	-0.276	0.757	0.000	1.000
Student race	0.107	1.003	0.000	0.793
EL	0.167	1.476	0.055	1.129
Suspension history	-0.279	0.718	0.000	1.000
Family income	0.135	0.900	0.077	0.872
Overall health	-0.088	1.096	0.168	0.839
Maternal education	0.106	1.154	0.134	1.196
Participated in at least one extracurricular activity	0.001	1.011	0.000	1.000
Frequency of reading together per week	0.043	0.937	0.000	1.000
Baseline measure				
Functional skills	-0.100	1.547	-0.071	1.408
Social skills	0.206	0.972	0.151	0.818
Living skills	0.340	0.913	0.055	0.640
Self-care skills	-0.035	1.222	0.236	0.968
Mental skills	0.118	1.697	0.090	1.533

Table 12. Standardized Mean Difference and Variance Ratio for Predictors Before and After Matching on Retention at Third Grade

Note. Functional, social, living, self-care and mental skills are baseline measures extracted from Wave 1. The standardized difference for covariate z between treatment (t) and control groups (t_0) are computed as $\delta_z(t) = [\widehat{\mu_z}(t) - \widehat{\mu_z}(t_0)] / \sqrt{\frac{\widehat{\sigma}_z(t)^2 + \widehat{\sigma}_z(t_0)^2}{2}}$. The variance ratio is computed as $\rho_z(t) = \{\hat{\sigma}_z^2(t)\}/\{\hat{\sigma}_z^2(t_0)\}$.

	Fourth grade			
	Before matching (N=105)		After matching (N=105)	
Child demographics	Standardized differences	Variance ratio	Standardized differences	Variance ratio
Gender	0.030	0.996	-0.039	1.021
Disability	-0.208	0.754	0.129	1.147
Student race	0.191	1.172	-0.168	0.562
EL	0.203	1.906	-0.065	0.836
Suspension history	0.000	1.015	-0.071	0.901
Family income	0.085	1.198	0.159	1.219
Overall health	-0.086	1.049	-0.058	1.025
Maternal education	0.208	1.365	-0.148	0.780
Participated in at least one extracurricular activity	-0.031	1.041	0.021	0.980
Frequency of reading together per week	-0.219	1.461	0.027	0.948
Baseline measure				
Functional skills	0.293	1.450	0.173	1.594
Social skills	-0.082	1.223	-0.130	0.863
Living skills	-0.263	1.197	-0.157	1.037
Self-care skills	-0.429	1.104	-0.268	0.754
Mental skills	-0.349	1.436	-0.196	1.650

Table 13. Standardized Mean Difference and Variance Ratio for Predictors Before and After Matching on Retention at Fourth Grade

Note. Functional, social, living, self-care and mental skills are baseline measures extracted from Wave 1. The standardized difference for covariate z between treatment (t) and control groups (t_0) are computed as $\delta_z(t) = [\widehat{\mu_z}(t) - \widehat{\mu_z}(t_0)] / \sqrt{\frac{\widehat{\sigma}_z(t)^2 + \widehat{\sigma}_z(t_0)^2}{2}}$. The variance ratio is computed as $\rho_z(t) = \{\hat{\sigma}_z^2(t)\}/\{\hat{\sigma}_z^2(t_0)\}$.

	Fifth grade			
	Before matching (N=133)		After matching (N=133)	
Child demographics	Standardized differences	Variance ratio	Standardized differences	Variance ratio
Gender	0.152	0.860	0.105	0.879
Disability	-0.217	0.764	-0.193	0.793
Student race	0.143	0.882	-0.011	0.656
EL	0.018	1.052	-0.246	0.515
Suspension history	0.105	1.072	0.031	1.017
Family income	-0.323	1.169	-0.018	1.552
Overall health	-0.036	1.030	0.113	0.916
Maternal education	0.123	0.829	0.233	1.358
Participated in at least one extracurricular activity	-0.199	1.152	0.232	0.808
Frequency of reading together per week	0.019	0.996	0.148	0.878
Baseline measure				
Functional skills	0.385	1.755	-0.048	1.177
Social skills	-0.292	1.039	-0.115	0.707
Living skills	0.279	1.210	-0.101	0.873
Self-care skills	-0.098	1.533	0.151	0.929
Mental skills	-0.379	1.674	0.053	1.150

Table 14. Standardized Mean Difference and Variance Ratio for Predictors Before and After Matching on Retention at Fifth Grade

Note. Functional, social, living, self-care and mental skills are baseline measures extracted from Wave 1. The standardized difference for covariate z between treatment (t) and control groups (t_0) are computed as $\delta_z(t) = [\widehat{\mu_z}(t) - \widehat{\mu_z}(t_0)] / \sqrt{\frac{\hat{\sigma}_z(t)^2 + \hat{\sigma}_z(t_0)^2}{2}}$. The variance ratio is computed as $\rho_z(t) = \{\hat{\sigma}_z^2(t)\}/\{\hat{\sigma}_z^2(t_0)\}$.

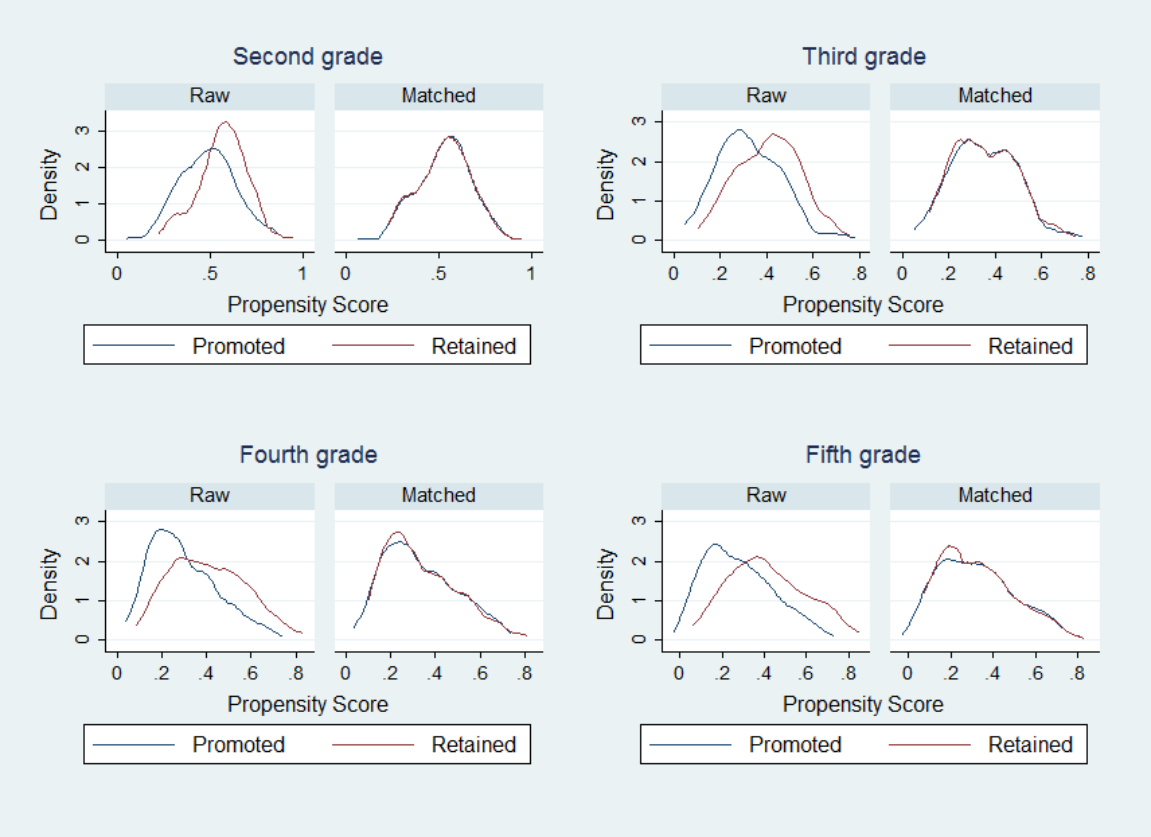

Figure 5: Density Plot of Retention Status at Second, Third, Fourth and Fifth Grade Before and After Matching

Hierarchical Linear Modeling (HLM) Predicting Academic Performance Using Matched Cases Identified From Second to Fifth Grade

Stacking all the matched cases from each grade level, a total of 484 matched students were identified with 182 students ever retained from second to fifth grade. Table 15 and 16 present the model results pertaining to the trajectories for the four academic subtests (letter word identification, passage comprehension, applied problem and calculation) among matched sample of students while controlling for the effect of retention, student race, EL status, and disability type. Two models were included for each of the academic subtests. Model 1 contains age, age squared and retention status. Model 2 adds three student level predictors (student race, disability types, and ELL status).

Letter word identification. Results predicting letter word identification using the matched sample of students in table 15 indicate that retention results in no significant difference in the outcome while controlling for the effects of student age. Model 2 further indicates while controlling for all the other covariates, retention status still produced no significant difference in the outcome. Instead, the change in the outcome is more related to the disability types as students with mental retardation ($\gamma_{04[MR]}$ = -36.96, p<.001) and other low incidence disabilities which include traumatic brain injury and multiple disabilities ($\gamma_{04[OLI]}$ = -33.88, p<.001) scored significantly lower on letter word identification than those with autism while holding all the other predictors constant.

Passage comprehension. Model 1 results in table 15 show that retained students scored no statistically different compared to their promoted peers on passage comprehension while controlling for the effect of student age. Model 2 indicates that once taking all the other predictors into account, retention still produced no significant differences in passage comprehension. Students with SLD, SLI, EBD, VI, OI, and OHI scored on average 13.85, 14.62, 15.02, 12.19, 10.64 and 10.49 points significantly higher

than their peers with autism on passage comprehension. By contrast, students with MR, and other low incidence disabilities scored on average 12.65 and 12.85 points significantly lower than their autism peers on passage comprehension while holding all the other predictors constant.

Applied problem. Model 1 results in table 16 indicate that retained students scored no statistically different than their promoted peers on applied problem while controlling for the effect of student age. Model 2 indicates retention results in no significant differences while controlling for all the other covariates. Instead, the change in the outcome is significantly related to student disability types. For instance, students with SLD ($\gamma_{04[SLD]}$ = 31.65, p<.001), SLI ($\gamma_{04[SLI]}$ = 36.35, p<.001), EBD ($\gamma_{04[EBD]}$ = 37.46, p<.001), HI ($\gamma_{04[HI]}$ = 18.20, p<.01), VI ($\gamma_{04[VI]}$ = 29.22, p<.001), OI($\gamma_{04[OI]}$ = 23.04, p<.001) and OHI ($\gamma_{04[OHI]}$ = 28.36, p<.001) scored significantly higher than their peers with autism on the outcome while holding all the other predictors constant.

Calculation. Model 1 results predicting calculation in table 16 indicate that retention results in an average 4.92 points significantly lower on calculation while controlling for the effect of student age. Model 2 indicates that while controlling for all the other covariates, retained students still scored 4.0 points significantly lower than their promoted peers. In addition, students with SLD, SLI and VI scored on average 9.38, 13.21 and 10.72 points significantly higher on calculation than their peers with autism. In contrast, students with mental retardation and other low incidence disabilities scored on average 15.83 and 13.12 points significantly lower than their counterparts with autism.

	Letter word identification		Passage comprehension	
	M1	M2	M1	M2
Age	7.558***	7.628***	3.789***	3.844***
	(0.443)	(0.442)	(0.296)	(0.297)
Age^2	-0.476***	-0.464***	-0.445***	-0.423***
	(0.124)	(0.123)	(0.0921)	(0.0919)
Retained	-5.360	-3.903	-2.928	-1.385
	(3.731)	(3.533)	(2.198)	(1.999)
Black		-0.480		-0.0510
		(4.609)		(2.699)
Hispanic		-0.329		3.758
		(5.500)		(3.723)
Asian and other		2.010		-1.883
		(13.47)		(8.895)
ELL		-1.458		-4.864
		(5.203)		(3.450)
SLD		-0.784		13.85***
		(5.508)		(3.023)
SLI		6.405		14.62***
		(6.641)		(4.171)
MR		-36.96***		-12.65*
		(8.032)		(4.934)

Table 15: Hierarchical Linear Regression Model Predicting Reading Performance Using Matched Sample of Students

	Letter word identification		Passage comprehension	
	M1	M2	M1	M2
EBD		8.331		15.02***
		(7.259)		(3.986)
HI		-8.617		3.366
		(6.627)		(4.345)
VI		6.059		12.19*
		(9.063)		(5.832)
OI		3.307		10.64**
		(6.908)		(4.076)
OHI		2.045		10.49**
		(6.209)		(3.580)
OLI		-33.88***		-12.85**
		(7.439)		(4.932)
Intercept	478.4***	484.4***	482.5***	477.6***
	(2.009)	(4.594)	(1.275)	(2.689)
Random effect				
Student	3.576***	3.475***	3.077***	2.953***
	(0.0388)	(0.0408)	(0.0394)	(0.0418)
Residual	2.664***	2.665***	2.464***	2.464***
	(0.0503)	(0.0502)	(0.0505)	(0.0501)
N	1158	1158	1164	1164

Note. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15, cont.

	Applied problem		Calculation	
	M1	M2	M3	M4
Age	5.396***	5.475***	4.922***	5.011***
	(0.438)	(0.428)	(0.333)	(0.327)
Age^2	-0.304*	-0.267	-0.482***	-0.470***
	(0.154)	(0.150)	(0.0989)	(0.0966)
Retained	-4.410	-2.355	-4.923*	-4.003*
	(3.316)	(2.996)	(1.936)	(1.806)
Black		-0.0442		1.787
		(3.012)		(1.883)
Hispanic		3.048		0.809
		(4.527)		(3.313)
Asian and other		6.812		4.646
		(10.21)		(5.596)
ELL		-4.307		2.958
		(4.560)		(3.088)
SLD		31.65***		9.378**
		(5.427)		(3.217)
SLI		36.35***		13.21***
		(6.668)		(3.594)
MR		-8.348		-15.83***
		(7.266)		(4.180)

Table 16: Hierarchical Linear Regression Model Predicting Math Performance Using Matched Sample of Students

	Applied problem		Calculation	
	M1	M2	M3	M4
EBD		37.46***		4.936
		(6.009)		(3.731)
HI		18.20**		2.193
		(6.059)		(4.039)
VI		29.22***		10.72*
		(8.039)		(4.256)
OI		23.04***		3.597
		(6.971)		(3.223)
OHI		28.36***		4.018
		(5.726)		(3.561)
OLI		2.136		-13.12**
		(7.296)		(4.746)
Intercept	481.6***	462.9***	499.4***	496.9***
	(2.020)	(5.065)	(1.229)	(2.861)
Random effect				
Student	3.452***	3.323***	2.933***	2.819***
	(0.0374)	(0.0428)	(0.0458)	(0.0523)
Residual	2.727***	2.726***	2.467***	2.466***
	(0.0562)	(0.0561)	(0.0320)	(0.0322)
N	1140	1140	1101	1101

Note. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 16, cont.

CHAPTER 5: DISCUSSION

The present study is the first to explore grade retention rates among students with disabilities using nationally representative data. The use of national level data enables researchers to examine the trend and effects of grade retention for a large group of sampled students with disabilities in a more consistent way as the state level policies regarding grade retention may vary. It is also one of the first large-scale studies to examine academic and behavioral trajectories among students with disabilities retained at different grade levels. The use of propensity score matching makes the present study the first to draw conclusion on the causal effect of retention on academic performances for students with disabilities. The current study specifically aimed to explore questions within elementary school setting as grade retention is more often seen at this stage in the United States (Hauser et al., 2006). In order to include as many students with disabilities as possible, the Special Education Elementary Longitudinal Study (SEELS) dataset was selected because it is by far the largest national-wide longitudinal dataset that exclusively contains students with disabilities who were followed along from age 7 until 17. As such, this study represents one of the most comprehensive studies on grade retention of a rarely explored student population that is at higher risk of being retained.

WORKING HYPOTHESES

Retention Rates

Findings of retention rates from this study offer partial support for the study hypotheses. As predicted, the retention rates among students with disabilities during elementary grades are much higher than the rates of students without disabilities reported in previous papers and reports. In the current study, more than 90% of students with disabilities were retained at least once. Once disaggregated by grade level, retention rates

are higher for students in early primary grades (i.e., kindergarten, first and second grade) than later grades and this is consistent with what Hauser and colleagues (2006) found in their study. In addition, the retention rates for students with disabilities remain higher than that of students without disabilities when disaggregated on grade levels from kindergarten to fifth grade. For instance, this study found retention rates for students with disabilities from kindergarten to fifth grade ranging from 7% to 43%, approximately six to seven times higher than rates reported in the study by Hauser and colleagues (2005). Such drastic differences could be explained by source of data, sampled population and how retention was defined in the two studies. Hauser and colleagues (2005) used Current Population Survey (CPS) as the data source which only includes population without disabilities. Sampled population in CPS are youth aged 5 to 20 from 1996 to 2003 whereas SEELS sampled students with disabilities aged 7 to 17 from 1999 to 2005. Last but not the least, Hauser and colleagues (2005) defined retention using the modal age of a grade as the proxy whereas the current study used direct measure to define retention.

Contradictory to previous research that identify moderate to large variations in retention rates among student demographic characteristics for students without disabilities, there is little variations in retention rates for students with disabilities in this study. The retention rates among students with disabilities vary slightly by gender, race, EL status, and free or reduced-priced lunch program. Consistent with previous findings that retention rates are higher among minority students, this study also identified slightly higher rates among Hispanic and Black students with disabilities although the difference is very small (Willson & Hughes, 2006). In addition, the present study, for the first time, reported retention rates by disability types. Retention rates are slightly higher for students with SLI (96%), MR (92%), visual impairment (92%) and SLD (90%) whereas lower for students with EBD (75%). This study also identified some patterns associated with retention rates

that are contradictory to previous empirical evidences in which rates are higher for students from poorer families (Hauser et al., 2006; Jimmerson et al., 1997; Mantzicopoulos, 2003; Tingle, Schoeneberger, & Algozzine, 2012). A possible explanation for the higher rates among students not enrolled in free lunch program may be a result of the measure used in SEELS to indicate family poverty. Although students from low-income families are more likely to be retained, the relationship between poverty and grade retention is further complicated by findings of non-significant association between poverty and grade retention (Beebe-Frankenberger, Bocian, MacMillan, & Gresham, 2004; Willson & Hughes, 2006). Moreover, recent researchers also start questioning the accuracy of using free lunch program enrollment as the indicator of family socioeconomic status as it may not accurately reflect student's access to economic resources (Harwell & LeBeau, 2010). Little variation of retention rates associated with gender, race, English language learner status and free or reduced-price lunch enrollment might also result from the characteristics of sampled population in SEELS. It was reported that "SEELS oversampled students in low-incidence disability categories to ensure equal precision in measuring all disability categories" (Wei, Blackorby, & Schiller, 2011, p.92). Consequently, about 91% of sampled children in SEELS started to have difficulty or disability condition before the age of 5 and 80% of them started receiving special services prior to their participation of SEELS. The oversampling makes SEELS an extreme case in which the majority of the sampled children were having severe cognitive or physical disabilities that almost all of them were retained regardless of their background demographics.

A glimpse of the descriptive statistics of reading and math performances by retention status at kindergarten to second grade shows retained students have slightly lower performances on the subtests of letter word identification, passage comprehension, applied problem and calculation. However, students retained at third grade have slightly higher

scores on the subtests of letter word identification and applied problems. By fourth grade, retained students slightly outperformed promoted students in letter word identification, applied problem and calculation. At fifth grade, retained students slightly outperformed their promoted peers on all four subtests. However, these findings should be interpreted with caution as the Wald means test shows the averages of student academic achievement (letter word identification, passage comprehension, applied problem and calculation) by retention status are not statistically different. In terms of social-behavioral outcome, retained students have slightly higher scores on the standardized composite social-behavioral outcome score than their promoted peers at kindergarten, second, and third grade whereas promoted students have slightly higher scores than retained students at first, fourth and fifth grade. However, the Wald means test also shows there is no statistical difference between the two groups on the composite social-behavioral outcome.

Academic and Behavioral Trajectory of Retained Students

Contradictory to the hypothesis on academic trajectories, students who were retained at kindergarten, first and second grade reveal no significant difference as opposed to their promoted peers on reading performance trajectories from age 7 to 17. Retention status only produced significant coefficients when predicting math performance at kindergarten. Specifically, students retained at kindergarten scored significantly lower than their promoted peers in calculation while controlling for all the other covariates. Comparisons of trajectories between retained and promoted students on reading performance at kindergarten, reading and math performance at first grade, reading and applied problem at second grade demonstrate no significant differences. This finding is consistent with previous studies in which retention does not benefit students academically (Jimerson, 2001; McCoy & Reynolds, 1999). The results also indicate that retention might

not only be an ineffective intervention in facilitating subsequent academic development given students retained did not score above their promoted peers while spending an extra year, but it also may hinder students from performing better, as reflected in the significant lower math scores among retained students.

The estimated behavioral trajectory also goes against with the hypotheses in that retained students were found to produce no significant difference on social-behavioral trajectories compared to their promoted peers at kindergarten, first and second grade. Unlike some of the previous studies that identified mixed (negative or positive) effects of retention on student social-behavioral outcome, this study concludes repeaters at kindergarten, first and second grade developed similar social-behavioral trajectories compared to their promoted peers (Cochan & Qadir, 2004; Jimerson & Ferguson, 2007; Stearns et al., 2007). The variation in findings across studies may reflect differences in measures, sample characteristics and analytical models. For instance, this study exclusively focused on students with disabilities, and unlike previous studies that examined a specific domain of student socioemotional aspect such as self-esteem, social adjustment or problem behaviors, the social-behavioral outcome used in the current study is a composite standardized score generated based on a list of teacher evaluation questions measuring different social-behavioral aspects such as self-regulation, ability to pay attention, follow directions, and peer relationships (Byrd, Weitzman & Auinger, 1997; Cochan & Qadir, 2004).

Although less relevant, the finding that disability types, gender, EL status and suspension history were significantly associated with social-behavioral outcomes suggests that student demographics and school experiences are playing more important roles when predicting their academic and social-behavioral performance. The findings are consistent with what Wei and colleagues (2011, 2012) found in their study when predicting student

reading and math performances using the same dataset while controlling for the effect of disability, gender, student race and family SES. The findings associated with disability types in predicting reading and math achievement are consistent with what Wei and colleagues reported in their studies. For instance, this study identified students with certain disability types such as those with SLI, EBD, VI, OHI had greater reading and math performances than the reference group (autism) whereas students with other types of disabilities such as MR and other low incidence disabilities (traumatic brain injury and multiple disabilities) scored below the reference group. In addition to disability types, this study also identified suspension history and home geography to be significant predictors of academic achievement. A noteworthy point from the results is the student age at the initial Wave and baseline measurement are consistently significant predictors. Older age at the beginning of the grade is associated with negative academic outcomes whereas higher scores on the baseline measurement are associated with positive academic outcomes over the 10-year span. This study also reveals that there was a deceleration in reading and math growth over time, indicating students with disabilities decelerated significantly slower and their reading and math growth trajectories flattened out as they moved onto high school. The findings indicate that more attention needs to be directed at those factors than retention since they are making more contributions to the changes in student reading and math performances.

Causal Retention Effect Using Propensity Score Matching

The findings of retention effect on student academic performances show that retained students produced similar reading performances as opposed to their promoted peers but retention resulted in a significant lower performance on the subtest of calculation. The negative effect of retention on mathematical skills still holds after controlling for other

variables such as student race, disability types and ELL status. The findings suggest that the effect of retention may be achievement domain-specific. Similar to what Wu and colleagues (2008) found in their study, retention results in significant negative effects on mathematical skills but had no significant effect on reading skills. Calculation serves as a foundation for later increasingly complicated and advanced mathematical skills given the hierarchical nature of mathematics skills (Peng et al., 2016). In the absence of retention, students with disabilities were found to scored 7.08 score points higher on calculation by one-year change in their ages and retention could delay the normal development of such fundamental mathematical skills, resulting in deteriorated performances in calculation compared to their promoted peers (Wei, Lenz, & Blackorby, 2012). When repeating a grade, it is reasonable to expect grade retention may produce positive effect on student's academic outcomes given the extra time provided for students to practice the necessary mathematic skills required to be successful. However, a retained student who struggles with mathematics often times also has difficulties in reading and he or she may be directed to spend more time at reading than math, which in turn might explain why retention is detrimental to math. The finding of a negative association between grade retention and math achievement is counter to previous literature that found a beneficial effect of retention (Hughes, Chen, Thommes, & Kwok, 2010). The variation in findings between this study and previously published studies could be attributed to the sample selection, the use of assessment, and policy context. First, this is the first study exclusively situated within students with disabilities. Although there are studies that focused on student subgroups such as those who are at higher risk for retention, or those with poor learning related skills, no former studies have focused their studies exclusively on students with disabilities. Hence, a change in criteria to select students in retention studies may contribute to changes in the effectiveness of retention (Hughes et al., 2010). In addition to the sample selection,

sample size after the matching procedure might also affect the findings. In this study, less than 150 students were matched for each grade level and subsequently used for later analyses whereas previous studies that found significant retention effects produced about 250 to 400 matched cases (Hughes, Chen, Thommes, & Kwok, 2010; Moser et al., 2012). Second, the effects of retention were found to differ based on the used achievement standard (i.e., curriculum-aligned test or nationally standardized test). Curriculum-aligned tests can generate grade equivalent scores, similar to same-grade comparisons used in the study by Hughes and colleagues whereas nationally standardized test generates age equivalent scores such as the Rasch-based W scores used in the current study (Hughes, Chen, Thommes, & Kwok, 2010). The effect of retention was less negative when using grade equivalent scores as the outcome as opposed to use age equivalent scores (Allen et al., 2009; Wu, West, & Hughes, 2008b). A third reason for results that diverge from previous studies is the SEELS data was collected at the time when the standards-based reform movement was started, and more emphasis was put on using high stakes testing for making decisions with regard to retention and move away from social promotion (U.S. Department of Education, 1999). Under such educational policy contexts, it is not surprising to see the majority of students with disabilities were retained as they were more likely to fall behind academically (Venable, 2015). However, we know little about the instructions those retained students were receiving while repeating the grade and the dataset does not include information on the kind of instructions and interventions provided for students with disabilities after they were retained. Grade retention by itself would not be expected to remediate the academic problems that students with disabilities experience and if they were retained without receiving specialized instructions, they may have less chance for success (Heubert & Hauser, 1999).

LIMITATIONS

This study contains several limitations that should be considered when summarizing the findings. To begin with, SEELS was not designed to include general education populations and thus a comparison of retention effect between students with and without disabilities is out of the question. Although the focus of the study is to examine the retention rates for students with disabilities, trajectories of students with disabilities retained at different grade, and retention effect on student academic outcomes, it also deserves the attention of future researchers to compare retention rates, performance gaps and retention effect between general and special education populations. Such comparisons could help further our understanding of how retention affects students with and without disabilities differently and what factors might cause the disparities.

Second, SEELS data was collected more than a decade ago, which makes it a less current up-to-date dataset. However, there is no recent data collected nationally wide that contains large number of students with disabilities as well as disaggregates by disability category. More recent national dataset such as the Early Childhood Longitudinal Study Kindergarten Class of 2010-2011 (ECLS-K:2011) contains some limitations which makes it an inappropriate database to use. First, the ECLS-K data only contains a small portion of students with disabilities (less than 10% of all students in the dataset are students with disabilities). Second, although the data collection still continues, the current data only covers students at elementary grades from kindergarten to fourth grade. More importantly, it does not collect data of retention status at student level.

Third, other variables not included in the logistic model to estimate propensity scores due to large number of missing values may also be predictive of retention status. In order to maximize the number of students after matching, the predictors included for the

logistic regression model to estimate propensity scores have to have less than 40% missing rate. However, some variables that were found to impact whether a student was retained captured in the dataset were not used because of the large percent of missing cases such as parental involvement at school (Willson & Hughes, 2006). In addition, there is also a lack of baseline measure for school readiness skill collected in SEELS (i.e., early academic skills in reading, math and general knowledge skills), which is another important predictor of grade retention (Davoudzadeh, McTernan, & Grimm, 2015).

Two other important limitations of this study relate to the methodology. First, although the propensity score matching provides control for differences between the retained and promoted groups on a list of baseline variables measured at the initial Wave, differences in some of these variables may occur after the first Wave. Therefore, it is extremely challenging to account for the effects of changes in measured covariates after the initial Wave of data collection (Rosenbaum & Rubin, 1983; Singer & Willet, 2003). The next limitation is related to the missing data on both the outcome and predictors. About 28% to 38% of students took the WJ-III at Wave 1, Wave 2 and 3, respectively. Although HLM model is flexible in handling missing values at level 1 and can substantially reduce bias due to missing data, it is not flexible in handling missing data at level 2 such as disability types, retention status or other demographic variables. In addition, about 21% of all students have valid data for the retention variable at each grade and the large number of missing value of retention status reduces the matched sample sizes by grade.

Last, although SEELS captured whether students were retained or not at each grade level, we know little about what instructions or interventions retained students were receiving during the retained year. In order to better explain why retention hinders students with disabilities from making progress, it is important to ask questions such as did students with disabilities get special education or additional support while being retained? Did they

receive individualized interventions? Is the retained year a simple repetition of the grade with the same teacher, same curriculum and same level of support as last year? However, such questions remain unknown in the dataset and should be explored in future studies. Moreover, alternatives to grade retention should also be considered in the future literature. Researchers have proposed a couple of alternatives besides retaining students if they do not meet predetermined academic standards such as offer summer school to students to catch up, provide before- and after-school programs and tutoring, train and hire qualified teachers to use a variety of teaching strategies to meet students' unique needs, or promote students with remedial instructions (Bolt, Krentz, & Thurlow, 2002; Darling-Hammond & Falk, 1997; Shepard & Smith, 1990). In one word, it makes more sense to critically investigate classroom and school practices when students are not making expected gains and pay more attention to the conditions under which students were thriving or failing. As Bowman (2005) pointed out, a critical question to ask is "whether or not students are unable to learn, choose not to learn, or do not have access to the resources they need to be successful learners" (p.45).

IMPLICATIONS FOR PRACTICE

With such a high proportion of students with disabilities retained at elementary grades and little to negative effect resulting from retention on student academic and social-behavioral outcome, it is less likely for teachers to recommend grade retention to students with disabilities. Moreover, considering the high financial cost associated with grade retention, results of the current study further add strong evidences for schools, teachers and policy makers to reconsider the use of grade retention and use the money more wisely (Bowman, 2005). This section discusses the present study's implications for practices from perspectives of schools, teachers and policy makers.

Implications for Schools

Although it may sound unrealistic to completely abandon grade retention practices in schools, it is important for schools and teachers to be aware of the retention rates and effects for students with disabilities and the early interventions that could help prevent retention from occurring. With so many students with disabilities at early grades (e.g., K-2) being retained, it raises concern on the progress of early school readiness for this particular group of students. Given early school readiness is a strong predictor of grade retention above child demographic variables such as race, family socioeconomic status, schools should develop more effective early reading programs to decrease the possibility of students being retained in early grades (Willson & Hughes, 2006). For students with disabilities who not only face challenges in academic sphere but also in behavioral aspect, it may also contribute to lowering retention rates if other interventions focusing on social and behavioral trainings were provided. In addition to early reading and behavioral programs, it was also suggested that preschool intervention programs can also help prevent academic failure. Such programs could target students with disabilities and assist them in developing early school readiness necessary for success at school. For instance, Head Start is an example of preschool intervention programs that has been proven to prevent retention from occurring once the targeted students were equipped with those skills (Barnett & Hustedt, 2005; Lincove & Painter, 2006).

Besides the preventive early interventions for grade retention, schools should also investigate the effectiveness of instructions students with disabilities were receiving while being retained. It is critical that schools' retention practices for students with disabilities align well with IDEA regulations. Essentially, students with disabilities who were retained should not be deprived of the opportunity to access the general curriculum so that they could still meet standards-based assessment and curriculum requirements (Bowman,

2005). Therefore, questions such as are retained students with disabilities receiving intensive remedial instructions while being retained should be asked frequently by school administrators to ensure that students with disabilities are not just repeating another year without any substantial changes or modifications on the instructions, resources and materials (Peterson & Hughes, 2011). Previous studies found retained students receiving supplemental and individualized instructions and supports made more gains in achievements (Stone & Engel, 2007). For students with disabilities, such intensive remedial instructions and practices may need to be sustained even beyond the retention year to ensure they could maintain their achievement.

Schools should also do a better job in publishing data pertaining to retention rates annually so that districts can be better informed of students' progress. Considering the heavy social and personal cost associated with grade retention, schools, districts and states should develop more efficient ways to keep track of how many students with disabilities were retained, who those retained students are and how retained students with disabilities performed compared to their promoted peers in both general and special education populations (Bowman, 2005).

Implications for Teachers

In order to enhance teachers' awareness of retention effect on student outcomes, teachers should be provided with more exposure to preparation and professional development programs and more opportunities to expand network with colleagues. For preservice teachers, it is critical for teacher preparation programs to help build and strengthen their awareness of the benefits and risks associated with grade retention. A well grasped knowledge of current research on retention could facilitate teachers in making better decisions about retaining students with disabilities. For example, it is suggested that

a wide range of factors should be taken into considerations by teachers when making decisions of retaining students with disabilities (Renaud, 2013). In addition to teacher preparation programs, preservice and in-service teachers should be involved in professional development programs and get better trainings in teaching strategies that will improve learning for students with disabilities (Bowman, 2005; Renaud, 2013). For example, training teachers to teach more effectively which include using assessment results to drive the subsequent instructions as well as monitoring students' progress using progress monitoring tools (Bowman, 2005). Progress monitoring tools such as Curriculum-Based Measurement (CBM) probes allow teachers to closely monitor students' progress on a regular basis so that teachers could identify less effective teaching practices and modify them accordingly if students were not making the expected progress (Shinn, 1998). It is also equally important for teachers to establish and maintain a healthy and open dialogue and communication with parents to discuss the possible reasons of academic failure, decisions to retain or not, and alternatives to retention that work best for students with disabilities. Previous studies found parental involvement in school is also an important predictor of retention (Willson & Hughes, 2006). By maintaining such conversations, teachers could not only encourage parents to be move involved in students' daily school life, but they could also get extra information on students' home literacy environment as well as how students with disabilities behave at home, all of which should be taken into considerations when making decision of retention. Last but not the least, results of the current study indicate no significant differences on retention rates by student demographic variables, but disability type is a significant predictor of academic and social-behavioral outcomes. Teachers should pay more attention to the disability profile of students when it comes to the failure to meet grade level expectations. Instead of repeating a grade, it may

be more worthwhile to understand the difficulties and struggles caused by students' disabilities.

Implications for Policy Makers

It has been almost 20 years since SEELS were collected. For students with disabilities in particular, efforts made by one stakeholder or another is far from enough. Rather, coordination from different facets of school reform, such as to align retention policies with Individuals with Disabilities Education Act (IDEA) regulations, should be guaranteed. An important prerequisite for policy makers to bear in mind is that students with disabilities should be given equal access to general curriculum even if they were retained so that they could meet standards-based assessment goals and requirements (Bowman, 2005). What also worth mentioning is the base on which decisions should be made regarding retention or promotion especially when it comes to students with disabilities. Should the decision of retaining or promoting a student with disabilities made only on the basis of a single assessment score? Should we also take into considerations of other factors such as disability type? These questions have to be answered more cautiously as the sole use of the assessment to make high stakes decision instead of taking into account other sources of evidences has been questioned more often by scholars and researchers (Thompson & Cunningham, 2000). Therefore, for students with disabilities, it is critical for policy makers to firstly guarantee their equal access to general curriculum to ensure teachers have the same expectations towards them as opposed to their peers without disabilities. More importantly, considerations on other sources of information and evidences on student academic and social-behavioral development should also be encouraged to take into consideration when students with disabilities fail academically or socially.

CONCLUSION

This dissertation study investigated retention rates, academic and behavioral trajectories of retained and promoted students, and retention effect on student academic outcomes exclusively for students with disabilities represented a comprehensive exploration of retention for students with disabilities for the first time. The results of this study indicate that retention rates for students with disabilities are much higher than that of all students or students without disabilities. The academic trajectories of retained students do not differ from that of their promoted peers except for math outcomes among students with disabilities who were retained at kindergarten and second grade, when retained students produced significantly lower trajectories. No statistical difference on social-behavioral trajectories were found between the retained and promoted groups. Retention effect on student academic outcome is primarily negligible with the exception of calculation. These results largely extend previous research on retention rates and effects. The current study analyzed growth trajectories of academic performance of students with disabilities on psychometrically strong measure of reading and math ability. Students with disabilities who were retained scored at no difference or at a significant lower level in academic and behavioral measurements as opposed to their promoted peers. In addition, the use of propensity score matching to control for the pre-treatment differences between the two groups provides strong evidence to date that grade retention fails to prove students with disabilities who are struggling could benefit from repeating an extra year.

Whereas prior studies of the grade retention effect on students' growth in achievement have assumed that retention may be a more appropriate intervention for students with the poorest academic and learning related skills, results from the current study challenge that assumption. Students with disabilities represent one particular student group that is characterized with poorer academic and behavioral skills, and retention does

not seem to benefit them in improving their academic and social-behavioral performances in the long term. Although the research evidence clearly demonstrates the ineffectiveness of grade retention, students with disabilities were retained as a result of their performance on state testing. Considering the cost associated with retention among students with disabilities could be high, more efforts should be directed to the exploration of other effective alternative interventions to prevent students with disabilities from lagging behind. Future research on the effects of retention among students with disabilities should take into account the instructions students were receiving while being retained and identify whether there are situations when grade retention produces beneficial outcome for students with disabilities. In addition, studies should also consider what other specific interventions can be provided to students with disabilities instead of grade retention.

Another noteworthy point about this study is that disability types were more salient predictors of student academic performances using propensity score matched samples after controlling for the effect of all the other factors such as student age and retention status. However, student race and EL status both fail to predict student academic outcomes. For instance, students with MR exhibited reading performance about 4.6 standard deviation whereas students with other low incidence disabilities scored about 4.71 standard deviation lower than their peers with autism on letter word identification. In contrast, students with SLD, SLI, EBD, VI, and OHI produced greater performances on passage comprehension than their peers with autism. Differences among disability categories are more evident in math growth trajectories. For instance, students with all disability types except for MR and OLI scored significantly higher on applied problem than their peers with autism. In terms of calculation, students with SLD, SLI and VI scored about 3 standard deviations higher than their peers with autism whereas students with mental retardation and other low incidence disabilities scored significantly lower than their peers with autism. The findings

point out the importance of providing students with disabilities, notably students with mental retardation and other low incidence disabilities, with more effective and timely interventions and remedial instructions before the performance gaps become even larger.

There are many ways grade retention could be improved and increased scrutiny around student outcomes should not be the only reason why we are paying closer attention to student success (Venable, 2015). Results of the current study provide some evidentiary support for grade retention theories that focus more on the environmental factors such as student background and school efforts. When students with disabilities fail at school, thorough considerations on specific challenges possibly resulting from disabilities, disadvantaged family conditions and inappropriate school practices that student, parents or schools face should be warranted. Unfortunately, the decision to retain students, particularly students with disabilities, was typically made without sufficient considerations on a more comprehensive view of students, which involves other important factors such as ability level, maturity, family environment and parental involvement. As a consequence, "the disappointing outcomes of retained students may well reflect the reasons they were held back in the first place rather than the consequences of being retained" (West, 2012, p.2). Retention is most likely to be successful if students who were retained were provided with additional research-based instructions or practices in subject areas where most of the effort was needed.

References

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. https://doi.org/10.3102/0162373709352239

Austin, P. C. (2009). Balance diagnosis for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, *28*, 3038–3107. https://doi.org/10.1002/sim

Barnett, W. S., & Hustedt, J. T. (2005). Head Start's lasting benefits. *Infants & Young Children*, *18*, 15–24.

Bashir, A. S., & Hook, P. E. (2009). Fluency: A key link between word identification and comprehension. *Speech, and Hearing Services in Schools*, *40*, 196–200.

Becker, H. S. (1963). *Outsiders*. New York: Free Press.

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.

Bianchi, S. M. (1984). Children's progress through school: A research note. *Sociology of Education*, *57*, 184–92.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*, 111–127.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328. https://doi.org/10.1080/19345740802400072

Bolt, S., Krentz, J., & Thurlow, M. (2002). *Are we there yet? Accountability for the performance of students with disabilities (Tech.Rep. No.33)*. Minneapolis, MN. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Technical33.htm

Bowman-Perrott, L., Benz, M. R., Hsu, H. Y., Kwok, O. M., Eisterhold, L. A., & Zhang, D. (2013). Patterns and predictors of disciplinary exclusion over time: An analysis of the SEELS national data set. *Journal of Emotional and Behavioral Disorders*, *21*(2), 83–96. https://doi.org/10.1177/1063426611407501

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

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*, 631–661.

Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1996). *Scales of Independent Behavior-Revised (SIBR)*. Rolling Meadows, IL: Riverside.

Byrd, R. S., Weitzman, M., & Auinger, P. (1997). Increased behavior problems associated with delayed school entry and delayed school progress. *Pediatrics*, *100*, 654–661.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Chen, Q., Hughes, J. N., & Kwok, O.-M. (2014). Differential growth trajectories for achievement among children retained in first grade. *The Elementary School Journal*, *114*(3), 327–353. https://doi.org/10.1086/674054

Cochan, B. I., & Qadir, S. A. (2004). Self-esteem of the repeaters: A mixed method study of elementary grade students. *Pakistan Journal of Psychological Research*, *28*(2), 277–296.

Cuddy, E. M., Frame, C., & DeVincentis, C. (1987). The effects of grade retention upon the social and psychological adjustment of elementary children. In *annual convention of the Association of Behavior Therapy*.

Darling-Hammond, L., & Falk, B. (1997). Using standards and assessments to support student learning. *Phi Delta Kappan*, *79*, 190–199.

Datar, A. (2006). Does delaying kindergarten entrance given children a head start? *Economics of Education Review*, *25*(1), 43–62.

Davoudzadeh, P., McTernan, M. L., & Grimm, K. J. (2015). Early school readiness predictors of grade retention from kindergarten through eighth grade: A multilevel discrete-time survival analysis approach. *Early Childhood Research Quarterly*, *32*(5), 183–192. https://doi.org/10.1016/j.ecresq.2015.04.005

Dawson, P. (1998). A primer on student retention: What the research says. *Communique*, *26*(8), 28–30.

Dougan, K., & Pijanowski, J. (2010). The effects of academic redshirting and relative age on student achievement. *International Journal of Educational Leadership Preparation*, 1–13.

Festinger, L. A. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology*, *88*, 3–17.

Frederick, C. B., & Hauser, R. M. (2008). Have we put an end to social promotion? Changes in school progress among children aged 6 to 17 from 1972 to 2005. *Demography*, *45*(3), 719–740. https://doi.org/10.2307/25475998

Gleason, K. a, Kwok, O.-M., & 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.

https://doi.org/10.1086/516667

Godard, P., McCracken, M., Rollin, J., Van Campen, J., Valdes, K., & Williamson, C. (2007). *Special education elementary longitudinal study data documentation and dictionary: Volume one*. Menlo Park, CA: SRI International.

Graue, M. E., & DiPerna, J. (2000). Redshirting and early retention: Who gets the "Gift of Time" and what are its outcomes? *American Educational Research Journal*, *37*(2), 509–534. https://doi.org/10.3102/00028312037002509

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher*, 39(2), 120-131. http://dx.doi.org/10.3102/0013189X10362578.

Hauser, R.M., Frederick, C.B., & Andrew, M. (2006). Grade retention in the age of accountability. In *Will Standards-Based Reform in Education Help Close the Poverty Gap?*

Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments answers from randomized experiments? *Psychological Methods*, *2*, 154–169.

Heubert, J. P., & Hauser, R. M. (1999a). *High stakes: Testing for tracking, promotion and graduation*. Washington, DC: National Academy Press.

Heubert, J. P., & Hauser, R. M. (1999b). *Short-term effects of grade*. Washington, DC: National Academy Press.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*(475), 901–910. https://doi.org/10.1198/016214506000000447

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

Individuals with Disabilities Education Act (2004).

Isenhart, L., & Bechard, S. (1987). *Dropout prevention. The ECS survey of state initiatives for youth at risk*.

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/10.2307/40645895

Jimerson, S.R., & Renshaw, T. L. (2012). Retention and social promotion. *Principal Leadership*, *13*(1), 12–16.

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. https://doi.org/10.1016/S0022-4405(99)00005-9

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, *30*(3), 420. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

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.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman test of educational achievement, second edition (KTEA-II)*. San Antonio, TX: Pearson Education.

Kerzner, R. L. (1982). *The effect of retention on achievement*. Union, NJ: Kean College of New Jersey.

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development*, *68*, 263–277.

Larsen, D.E., & Akmal, T. T. (2007). Making decisions in the dark:

Disconnects between retention research and middle-level practice. *NASSP Bulletin*, *91*(1), 33–56.

Lemert, E. (1967). *Human deviance, social problems and social control*. Englewood Cliffs, NJ: Prentice-Hall.

Lin, H., Freeman, L. S., & Chu, K. (2009). The impact of kindergarten enrollment age on academic performance through kindergarten to fifth grade. *European Journal of Social Sciences*, *10*(1), 45–54.

Lincove, J. A., & Painter, G. (2006). Does the age that children start kindergarten matter? Evidence of long-term educational and social outcomes. *Educational Evaluation and Policy Analysis*, *28*, 153–179. https://doi.org/10.3102/01623737028002153

Lorence, J., & Dworkin, A. G. (2006). Elementary grade retention in Texas and reading achievement among racial groups: 1994-2002. *Review of Policy Research*, *23*(5), 999–1033. https://doi.org/10.1111/j.1541-1338.2006.00247.x

Mantzicopoulos, P. Y. (1997). Do certain groups of children profit from early retention? A follow-up study of kindergartners with attention problems. *Psychology in the Schools*, *34*(2), 115–127. https://doi.org/10.1002/(SICI)1520-6807(199704)34:2<115::AID-PITS5>3.0.CO;2-P

Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, *101*, 101–14.

Martinez, B., & Vandergrift, J. A. (1991). *Failing students: Is it worth the cost?* Tempe: Morrison Institute for Public Policy. Arizona State University.

Mather, N., Schrank, F. A., & Woodcock, R. W. (2007). *Examiner's Manual: Woodcock-Johnson III Tests of Achievement Form C/Brief Battery*. Rolling Meadows, IL: Riverside.

McCoy, A., & Renolds, A. (1999). Grade retention and school performance. *Journal of School Psychology*, *37*(3), 275–295.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock-Johnson III*. Itasca, IL: Riverside.

McLeskey, J., & Grizzle, K. L. (1992). Grade retention rates among students with learning disabilities. *Except Child*, *58*(6), 548–554. https://doi.org/10.1177/001440299205800609

Meisels, S. J.(2002). Assessing readiness. In R.C. Pianta & M. J. Pianta (Eds). *The transition to kindergarten*. London: National Center for Early Development and Learning.

Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly*, *31*, 187–198. https://doi.org/doi: 10.2307/25474651

Morrison, F. J., Griffith, E. M., & Alberts, D. M. (1997). Nature-Nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*, 254–262.

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology*, *104*(3), 603–621. https://doi.org/10.1037/a0027571

NAEP Report Card. (2017). U.S. Department of Education, Institute of Education Sciences, National Centerr for Education Statistics.

Nikalson, L. B. (1984). Nonpromotion: A psycho-scientific solution. *Psychology in the Schools*, *21*, 485–499.

Oshima, T. C., & Domaleski, C. S. (2006). Academic performance gap between summer-birthday and fall-birthday children in grades K-8. *Journal of Educational Research*, *99*(4), 212–217.

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral

development. *Development and Psychopathology*, *13*, 297–315.

Peng, P., Namkung, J. M., Fuchs, D., Fuchs, L. S., Patton, S., Yen, L., … Hamlett, C. (2016). A longitudinal study on predictors of early calculation development among young children at risk for learning difficulties. *Journal of Experimental Child Psychology*, *152*, 221–241. https://doi.org/10.1016/j.jecp.2016.07.017

Peterson, L. S., & Hughes, J. N. (2011). Differences between retained and promoted children in educational services received prior to and after retention year. *Psychology in the Schools*, *48*, 156–165. https://doi.org/10.1001/pits.20534

Peterson, S. E., DeGracie, J. S., & Ayabe, C. R. (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal*, *24*(1), 107–118.

Planty,M., Hussar, W., Snyder, T., Kena, G., KewalRamani, A., Kemp, J., & Dinkes, R. (2009). The condition of education 2009 (NCES 2009-081). Washington, DC: U.S. Department of Education, Institute of Education Science, National Center for Education Statistics.

Plummer, D. L., & Graziano, W. G. (1987). Impact of grade retention on the social development of elementary school children. *Developmental Psychology*, *23*(2), 267–275.

Rachal, J., & Hoffman, L. M. (1985). *The effects of remediation and retention upon basic skills performance among elementary students participating in a state basics skills test program*. Baton Rouge, LA: Louisiana State Department of Education.

Range, C., Dougan, K., & Pijanowski, K. (2011). Rethinking grade retention and academic redshirting: Helping school administrators make sense of what works. *International Journal of Educational Leadership Preparation*, *6*(2), 1–12.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models application and data analysis methods*. Thousand Oaks, CA: Sage.

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

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago ' s high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, *27*(4), 309–340.

Rosenbaum, P.R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rosenbaum, P. R. (2002). Attributing effects to treatment in matched observational studies. *Journal of the American Statistical Association*, *97*, 183–192.

Rust, J., & Wallace, K. (1993). Effects of grade level retention for four years. *Journal of Instructional Psychology*, *20*(2), 162–166.

Schrank, F.A., McGrew, K.S., & Woodcock, R.W.(2001). *Technical Abstract* (Woodcock-Johnson III Assessment Service Bulletin No.2). Itasca, IL: Riverside Publishing

Scott, N. M. (2011). Please, not another push to get tough on student retention. *Planning and Changing*, *42*(3–4), 209–223.

Shadish, W. R., Luellen, J. K., & Clark, M. H. (2006). Propensity scores and quasi-experiments: A testimony to the practical side of Lee Sechrest. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 143–157). Washington, DC: American Psychological Association.

Shepard, L. A., & Smith, M. L. (1989). *Flunking grades: Research and policies on retention*. Philadelphia, PA: Falmer.

Shepard, L., & Smith, M. (1990). Synthesis of research on grade retention. *Educational Leadership*, *47*(8), 84–88.

Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: The Guilford 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 http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-01833-010&site=ehost-live bsilberglitt@scred.k12.mu.us

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smith, M. L., & Shepard, L. A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. *American Educational Research Journal*, *25*, 307–333.

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*, 210–240.

Stone, S., & Engel, M. (2007). Same old, same old? Students' experiences of grade retention under Chicago's ending social promotion policy. *American Journal of Education*, *113*, 605–634. https://doi.org/10.1086/518490

Texas Education Agency. (2012). Texas Education Agency academic excellence indicator system.

Tingle, L. R., Schoeneberger, J., & Algozzine, B. (2012). Does Grade Retention Make a Difference? *The Clearing House*, *85*(5), 179–185. https://doi.org/10.1080/00098655.2012.679325

Thompson, C. L., & Cunningham, E. K. (2000). Retention and social promotion: Research and implications for policy. ERIC Digest Number 161.

U.S. Department of Education. (1999). *Taking responsibility for ending social promotion*. Retrieved March 23, 2018, from http://www.ed.gov/PDFDocs/socialprom.pdf

Venable, S. (2015). Grade-level retention : Not always a positive strategy. *National Teacher Education Journal*, *8*(3), 55–60. Retrieved from http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=6111c18e-388c-4a0e-8acc-88bb512d83cf%40sessionmgr4010&vid=1&hid=4107

Wagner, M., Newman, L., Cameto, R., Levine, P., & Garza, N. (2006). *Overview of findings from Wave 1 of the Special Education Elementary Longitudinal Study (SEELS)*. Menlo Park, CA.

Walker, E. M., & Madhere, S. (1987). Multiple retentions: Some consequences for the cognitive and affective maturation of minority students. *Urban Education*, *22*(1), 85–102.

Walters, D. M., & Borgers, S. B. (1995). Student retention: Is it effective? *The School Counselor*, *42*(4), 300–310.

Warren, J. R. & Saliba, J. (2012). First- through eighth-grade retention rates for all 50 states: A new method and initial results. *Educational Researcher*, *41*(8), 320–329. https://doi.org/10.3102/0013189X12457813

Warren, J. R., Hoffman, E., & Andrew, M. (2014). Patterns and trends in grade retention rates in the United States, 1995-2010. *Educational Researcher*, *43*(9), 433–443. https://doi.org/10.3102/0013189X14563599

Wechsler, D. (2001). *Wechsler individual achievement test-second edition (WIAT-II_*. San Antonio, TX: Pearson Education.

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*(3), 154–165. https://doi.org/10.1177/0741932512448253

West, M.R. (2012). *Is retaining students in the early grades self-defeating?* Washington D.C.: The Brookings Institution

Willson, V. L., & Hughes, J. N. (2006). Retention of Hispanic/Latino

students in first grade: Child, parent, teacher, school, and peer predictors. *Journal of School Psychology*, *44*(1), 31–49. https://doi.org/10.1016/j.jsp.2005.12.001

Winsler, A., Hutchison, L. A., De Feyter, J. J., Manfra, L., Bleiker, C., Hartman, S. C., & Levitt, J. (2012). Child, family, and childcare predictors of delayed school entry and kindergarten retention among linguistically and ethnically diverse children. *Developmental Psychology*, *48*(5), 1299–1314. https://doi.org/10.1037/a0026985

Wu, W., West, S. G., & Hughes, J. N. (2008). Short-term effects of grade retention on the growth rate of Woodcock-Johnson III broad math and reading scores. *Journal of School Psychology*, *46*(1), 85–105. https://doi.org/10.1016/j.jsp.2007.01.003

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology*, *102*(1), 135–152. https://doi.org/10.1037/a0016664

Yamamoto, K., & Byrnes, D. (1987). Primary children's ratings of the stressfulness of experience. *Journal of Research in Childhood Education*, *2*(2), 117–121.